

**ESTIMATING THE EFFECT OF VEHICLE SPEEDS ON BICYCLE AND
PEDESTRIAN SAFETY ON THE GEORGIA ARTERIAL ROADWAY NETWORK**

A Thesis
Presented to
The Academic Faculty

By

Daniel F. Arias

In Partial Fulfillment
Of the Requirements for the Degree
Master of Science in Civil & Environmental Engineering,
Master of City & Regional Planning

Georgia Institute of Technology
August 2020

Copyright © Daniel F. Arias 2020

**ESTIMATING THE EFFECT OF VEHICLE SPEEDS ON BICYCLE AND
PEDESTRIAN SAFETY ON THE GEORGIA ARTERIAL ROADWAY NETWORK**

Approved by:

Dr. Kari Watkins
School of Civil & Environmental Engineering
Georgia Institute of Technology

Dr. Michael Hunter
School of Civil & Environmental Engineering
Georgia Institute of Technology

Dr. Michael Rodgers
School of Civil & Environmental Engineering
Georgia Institute of Technology

Date Approved: June 27, 2020

ACKNOWLEDGEMENTS

First, I owe my thanks to the great city of Philadelphia for, the first time in my life, offering resistance to my default setting of driving and parking everywhere I go. Thank you for forcing me to think about my options for transportation. And for opening my eyes to what a place can be if we make that concession. And yes, even for all the parking tickets.

To Dr. Kari Watkins, thank you for placing a bet on my success. It has been a privilege to work with you, learn from you, and catch your passion. It is contagious. I hope I can pass it on as I progress through my career.

To David Ederer, thank you for all the time you gave me, your patient teaching, and, perhaps most of all, for the gallons of coffee you provided me. You were my greatest influence during my few years at Georgia Tech. It is an honor to call you my friend.

Dr. Rodgers and Dr. Hunter, thank you for your careful guidance during this project, and for challenging me to meet your high expectations. I believe I grew much more because of it.

Thank you to the Georgia Department of Transportation, and especially Sam Harris, for providing this research opportunity. More importantly, thank you for your passion for the safety of vulnerable road users and your willingness to explore new avenues for ensuring that safety. Not only is it the reason why this thesis was possible, but it makes me hopeful for the future of my home state of Georgia.

Thank you to Meredith, my wife, for simply being there for me as I worked on this project. I hope I can do the same for you in the near future. Thanks to Danny Anderson for listening to my complaints, to Sam Kilkenny for whiffleball breaks, and to Cooper for being such a good boy.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES	vii
SUMMARY	viii
CHAPTER 1. Introduction	1
CHAPTER 2. Literature Review	4
2.1. Vehicle speeds and crashes involving all vehicle types	4
2.1.1. Vehicle speed and crash probability	4
2.1.2. Variation in vehicle speed and crashes	7
2.2. Factors in bicycle and pedestrian crashes	8
2.2.1. Vehicle speed and bicycle/pedestrian crashes	8
2.2.2. Crash count modeling techniques	14
2.2.3. Built environment factors and crash probability	16
2.2.4. Socioeconomic and sociodemographic characteristics	23
2.3. Summary	24
CHAPTER 3. Data	25
3.1. Data sources	25
3.1.1. Crash data	25
3.1.2. Speed and roadway data	26
3.1.3. Speed limit data	27
3.1.4. Climate data	27
3.2. Data preparation	28
3.2.1. Crashes: backfilling invalid spatial data	28
3.2.2. Identifying bicycle and pedestrian crashes	29
3.2.3. Processing TMCs and speeds	32
3.2.4. Conflating speed limits to TMCs	35
3.2.5. Conflating crashes to TMCs	37

3.2.6. Conflating ecoregions to TMCs	38
CHAPTER 4. Methodology	40
4.1. Model specification	40
4.2. Other included covariates	42
CHAPTER 5. Results	46
5.1. Speed percentiles and crash frequency	46
5.2. Speed differences and crash frequency	46
5.3. Other covariates	47
CHAPTER 6. Discussion	50
6.1. Speed differences	50
6.2. Speed percentiles	53
6.3. Posted speed	54
6.4. Other covariates	55
6.5. Limitations	56
6.6. Policy implications	57
6.7. Summary	61
CHAPTER 7. Conclusion	63
REFERENCES	67

LIST OF TABLES

Table 1: Included TMCs by Functional Class	32
Table 2: Results of posted speed limit conflation by tolerance distance	36
Table 3: Included TMCs by Ecoregion	38
Table 4: Summary statistics for each TMC	39
Table 5: Correlation matrix for included TMC variables	45
Table 6: Statewide Bike/Ped Crashes and Vehicle Percentile Speeds	48
Table 7: Statewide Bike/Ped Crashes and Vehicle Speed Distribution	49

LIST OF FIGURES

Figure 1: Example data where speed safety relationship is negative if functional classes are modeled together, or positive if modeled separately	5
Figure 2: The Solomon Curve	6
Figure 3: The Power Model	7
Figure 4: Taylor's conceptual model for speed and crash frequency	8
Figure 5: The S-shaped severe injury and fatality risk curves	12
Figure 6: Area-level social crash cost in Manhattan	13
Figure 7: Weighted nearest neighbor on nearby road links	15
Figure 8: Sub-linear growth rate between pedestrian activity and pedestrian crash counts	21
Figure 9: Example of Internal, External TMCs	27
Figure 10: Steps in data preparation and their inputs	28
Figure 11: Bicycle and pedestrian crashes cluster with population in Georgia	31
Figure 12: Non-primary TMCs overlap with primary TMCs	33
Figure 13: Included arterial roadways in Georgia	34
Figure 14: Illustration of Intersect process and resulting shapes and data attribution	36
Figure 15: Estimated marginal effects of 85th - 50th percentile speeds on crashes per 10 miles	52

SUMMARY

Despite a decreasing trend in overall crashes, bicyclist and pedestrian fatalities have increased steadily since 2009 in the United States (Cicchino & Hu 2016). A large body of research suggests vehicle speeds are a key contributing factor for crashes (Elvik et al. 2019). Furthermore, vehicle impact speed has been identified as the principal determinant of severity and death in the event of a pedestrian crash (Tefft 2013). However, there have been few studies of bicycle or pedestrian crash probability that incorporate detailed vehicle speed data. Newly available probe vehicle data in the state of Georgia makes it possible to study the relationship between bicycle and pedestrian crashes and speed across the network of Georgia arterial roadways.

The analysis uses INRIX® speed data and the Georgia DOT crash database and relates these data in a Negative Binomial crash count model for the year 2017. Models using speed percentiles (85th, 50th and 15th) and models using speed differences (85th - 50th and 50th - 15th percentile) are compared. A small set of covariates are included.

This study shows that the high speed difference (85th - 50th percentile) is a robust indicator of bicycle and pedestrian crash frequency on Georgia arterial roadways. The high speed difference outperformed the low speed difference (50th - 15th percentile), suggesting that the high end of the distribution is more important to crash prediction than the low end. Additionally, speed percentile models showed no clear, intuitive relationship to bicycle and pedestrian crashes.

In light of these results, planners and policymakers should identify arterial roadways with

high speeds, high spread of speeds at the top end of the distribution, and high bicyclists and pedestrian activity. To do so, a complete bicycle and pedestrian count data collection effort is needed. These target roadways should be considered for treatments which prioritize the reduction of the fastest speeds and limitation of exposure for unprotected road users. Finally, the practice of setting the speed limit at the 85th percentile speed (NTSB 2017) should end. Road user safety must supplant vehicle throughput and access to create a sustainable, equitable and just transportation system in Georgia.

Chapter 1

Introduction

Each year, roughly 40,000 deaths occur due to traffic collisions on American roadways. According to the Insurance Institute for Highway Safety (IIHS), traffic death tolls have begun to climb steadily after decades of year-over-year reductions (IIHS Yearly Snapshot 2019). Bicycle and pedestrian deaths are no exception. Pedestrian fatalities have steadily risen nationwide, by nearly 50% since 2009 (Cicchino & Hu, 2018). Bicycle fatalities have also increased steadily since 2009 (IIHS Bicyclists 2019). While the majority of traffic collisions, injuries, and deaths occur in vehicle-to-vehicle crashes, bicycle and pedestrian crashes are persistent and increasingly-common events in America.

In contrast to vehicle users, bicyclists and pedestrians have virtually no protection from impact in a crash event. In 2015, despite much lower mode share, 15% of total traffic fatalities were pedestrians, and 2.3% were bicyclists, according to the National Highway Traffic Safety Administration. Injury and fatality rates among bicycle and pedestrian crashes have risen steadily since 2006 (NHTSA 2017). Thus, bicyclists and pedestrians bear a disproportionate burden of injuries and deaths in traffic collisions. These elevated risks deter bicycle and pedestrian activity, impeding healthier and more environmentally sustainable lifestyle choices (Geyer et al 2006).

Vehicle speeds present significant risk to active transportation road users. It is well-documented that increases in vehicle speeds lead to greater risk of injury or death to bicyclists/pedestrians in the event of a crash (Tefft 2013). Put simply, a higher vehicle speed transfers more energy into a collision than a slower-moving vehicle (Aarts and van Schagen

2006). Tefft (2013) showed that at speeds less than 40 mph the likelihood of pedestrian death increases exponentially as the speed of the colliding vehicle increases.

However, there is evidence that vehicle speeds affect the likelihood of a crash, in addition to its severity. In theory, higher speeds reduce reaction time of parties when a conflict occurs, increasing the likelihood of a crash. Elvik et al. (2019) showed in a meta-analysis of speed and safety studies that crash frequency increases with speed increases according to the power model.

This study aims to estimate the risk that vehicle speeds pose to bicyclists and pedestrians using newly-available data sources. Using vehicle probe speed data, this study conducts a state-wide evaluation of Georgia roadways and their crashes involving active transportation users. The relationship between vehicle speeds and bicycle/pedestrian crashes/injuries is identified using a Negative Binomial regression model.

The availability of new data makes it possible to study a large portion of the Georgia arterial network in this study. Probe vehicle speed data from INRIX® enables the broad inclusion of arterial roadways around the state. This study is the first study of its kind to examine the relationship between speed and crashes involving bicyclists and pedestrians for the full arterial network in Georgia.

This study evaluates the explanatory power of operational vehicle speeds (especially the 85th percentile speed) in comparison with distributions of vehicle speeds. A critical hypothesis of this study is that absolute vehicle speed is less effective than the distribution of vehicle speeds in describing the likelihood of a vehicle-pedestrian or vehicle-bicycle crash. Intuitively, a wider distribution of vehicle speeds describes a more complex roadway environment, where slower and faster vehicles come into conflict. Examples include vehicles turning from driveways into traffic,

congestion accumulating, and congestion dissipating. In these environments, vehicle speed distributions are necessarily broader, and drivers must stay especially attentive to other vehicles in order to avoid crashes. It may be that in these environments, bicyclists and pedestrians are noticed later, and crashes involving these road users are more likely.

Even when controlling for other roadway and environmental characteristics, the high end of the vehicle speed distribution is a robust and significant indicator of vehicle-pedestrian and vehicle-bicyclist crash risk. A high speed difference, defined in this study as the difference between 85th and 50th percentile speeds, is a more effective indicator of crash risk than absolute speed percentile values. The findings in this study suggest that planners and engineers should identify areas with large speed distributions, especially at high vehicle speeds, and work to reduce the fastest speeds on these roadways.

Chapter 2

Literature Review

The review of literature breaks out into two broad parts. The first section examines the extensive research on the relationship between vehicles speeds and frequency of crashes of any kind. The second section delves into bicycle and pedestrian crashes, examining the ways in which these crashes are generally studied, in reference to speed as well as other factors believed to contribute to bicycle and pedestrian crash incidence.

2.1. Vehicle speeds and crashes involving all vehicle types

2.1.1 Vehicle speed and crash probability

The relationship between crashes and vehicle speed is frequently studied. While there is no consensus on the nature of the relationship between speed and crash frequency (Aarts and Van Schagen 2006), many studies assume that a positive relationship exists (Silvano and Bang 2016). Intuitively, crash probability would increase with speed, as available reaction time to conflicts is reduced and avoidance maneuvers are more challenging (AASHTO, 2010). However, there are many complicating factors that obscure the nature of the relationship.

Historically, some studies have found that the relationship between vehicle speed and crash frequency is negative; that is, higher speeds may reduce crash rates. These results are usually explained by important confounding factors, like an urban context, intersection density, and congestion (Aarts and van Schagen 2006). Elvik et al. (2004) also mentioned that the ecological fallacy is a factor in the speed-safety relationship: when all functional classes are

modeled together, the relationship takes on a different shape than if the functional classes are accounted for (see Figure 1). Characteristics of functional classes, especially density of access points and population density, drive up crash rates for lower functional classes, which also happen to have lower speeds. More recent studies, described below, tend to control for these factors and generally have not found a negative speed-safety relationship.

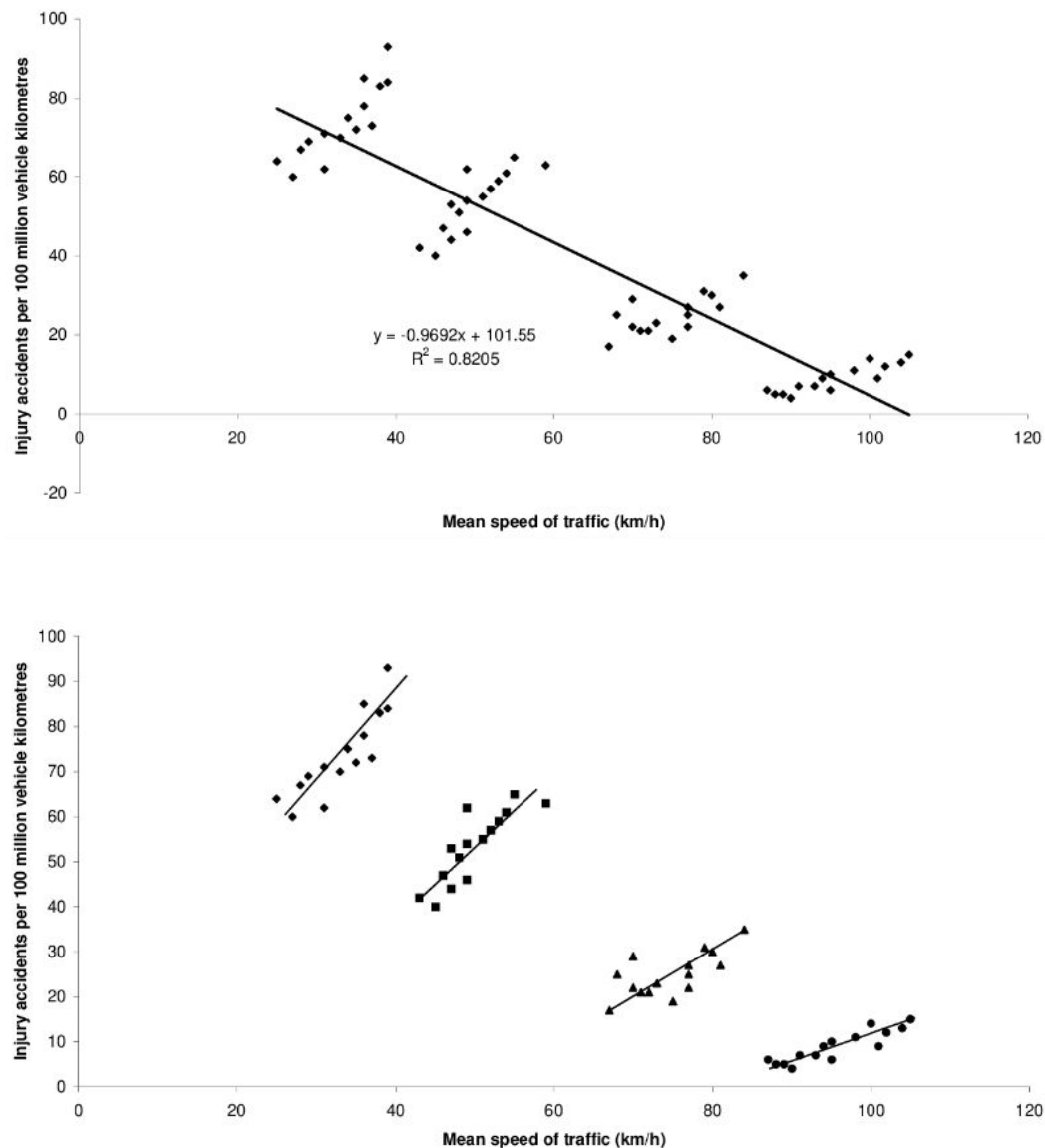


Figure 1: Example data where the speed safety relationship is negative if functional classes are modeled together (top) or positive if modeled separately (bottom) (Elvik et al. 2004).

Perhaps the most famous description of the speed-safety relationship is from Solomon (1964) who said that extreme speeds, both low and high, are associated with higher crash rates than moderate speeds. The resulting U-shaped curve (Figure 2) has become central to transportation planning since its inception, implying that there is an “ideal” speed at which roads are safest. This study has also been challenged in recent years (Aarts and van Schagen, 2006), especially on the grounds that congestion, which Solomon did not consider in his model, accounts for the increased crash rates at low speeds.

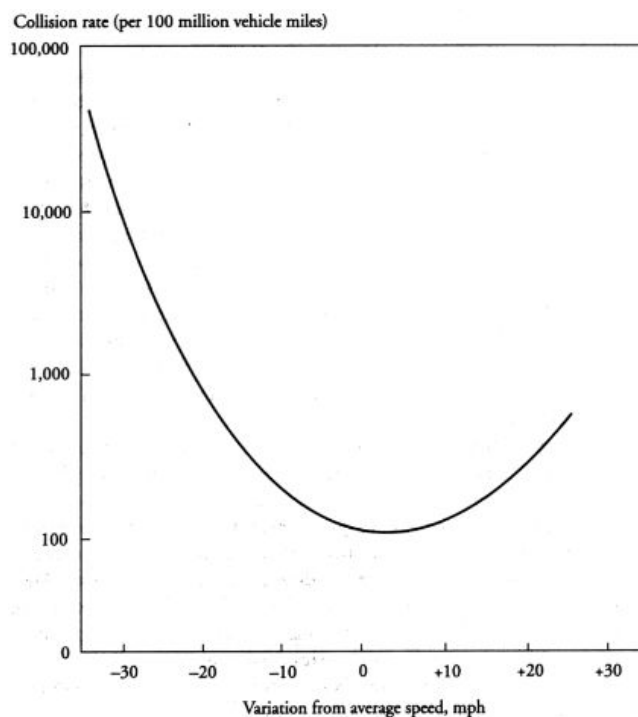


Figure 2: The Solomon Curve (1964) suggests crash rate is higher at lower and higher speeds.

Other studies have found a strictly positive relationship between speed and crashes. Elvik et al. (2004, 2019) has shown through meta-analyses of speed and crash rates that the

relationship follows a power model. That is, not only do speed increases always increase crash rates, but a speed increase at higher speeds has a greater effect on crash rates than the same increase at low speeds (Figure 3). The power model is commonly used as the basis for studies that estimate speed changes and wish to extrapolate those changes to an impact on safety (Silvano & Bang, 2016).

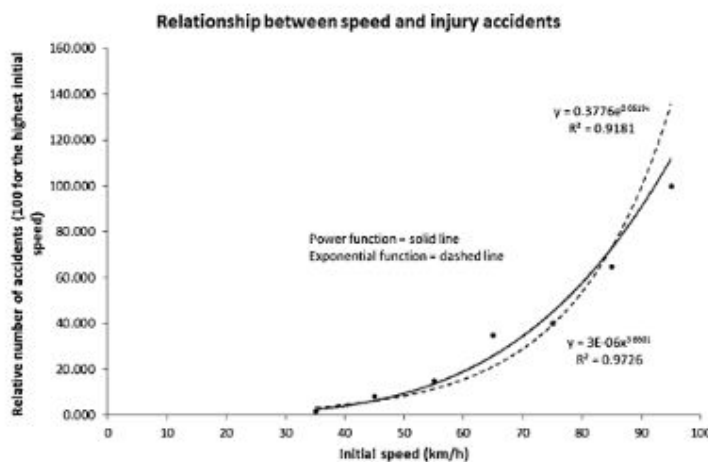


Figure 3: The Power Model estimation of vehicle speeds and crash rates (Elvik 2019)

2.1.2. Variation in vehicle speed and crashes

Speed variation is the differences in vehicle speeds on a stretch of roadway. Greater spread of vehicle speeds is generally considered to increase crash probability, since large differences in speed create conflict between vehicles who pass each other more frequently. Research has consistently provided evidence to support this intuition (Taylor 2000, Wang et al. 2018).

It should be noted that the details of measuring speed variation are important to drawing conclusions about the speed and safety relationship. A conceptual model is shown in Figure 4.

There are many ways to measure the distribution of speeds. Some commonly-used metrics, e.g. standard deviation of vehicle speeds, consider the low end and high end of speed distribution equally, but metrics like percentage of speeders recognize the outsized impact of the high end of the distribution (Taylor 2000). Little research has formally compared different measures of speed variation as explanatory factors in crash rates.

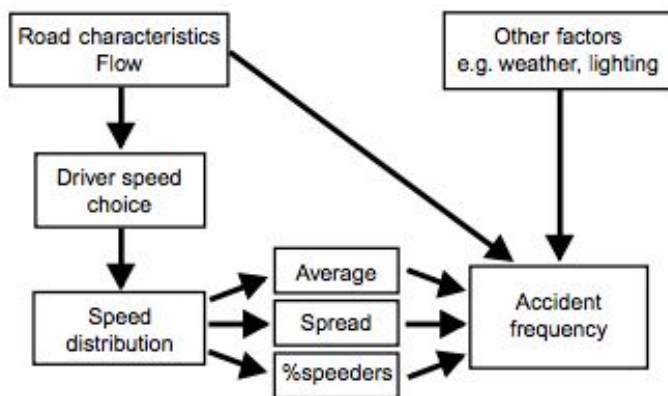


Figure 4: Taylor's conceptual model for speed and crash frequency. Spread of speeds and percent of speeders consider the whole speed distribution with emphasis on the high end.

2.2. Factors in bicycle and pedestrian crashes

2.2.1. Vehicle speed and bicycle/pedestrian crashes

The vehicle speed and bicycle/pedestrian crash literature breaks out broadly into two categories, both of which are described in detail below. The first category examines crash frequency or crash rate in an attempt to understand which factors influence the likelihood of pedestrian crashes occurring. Speed is usually incidental to the modeling specification and interpretation of results in these studies, since area level speed data has historically been difficult

to obtain and the nature of the relationship between speed and crash likelihood is not universally agreed-upon.

The second line of study assumes a crash has already occurred and attempts to measure the relationship between crash characteristics and the likelihood of more severe injuries or fatalities. Speed is often critical to these studies, both because fundamental physics ensures a relationship between the two and because speed data in the event of a crash is more broadly available. Speed estimation in crash reports has been standard practice for some time, and black box recording devices in vehicles have become popular more recently.

2.2.1.1. Crash frequency versus crash rate

Crash frequency is the count of crashes in a specified time and space. This is a common dependent variable in pedestrian and bicycle safety analyses. Data on crash counts is frequently available from public sources, and modeling techniques in crash counts are robust and well established, as described in the section below.

Crash rate is most commonly expressed as the number of crashes as a ratio of population in a specified time and space. The key advantage of crash rate over frequency is its description of the level of risk associated with transportation in a certain mode; crash rate describes how likely an individual in an area is to be involved in a crash.

Models of pedestrian crash frequency should account for pedestrian and vehicle activity activity, just as models of vehicle crash frequency should account for vehicle miles traveled (Geyer et al. 2006). Interpretation of model coefficients should also account for this difference. A model might show that bicycle crashes are more common in a city center, but more bicycling

activity may be expected in the city center, so this finding only indicates greater bicycle crash risk if the researcher controlled for bicycle distance traveled (Kaplan & Prato 2015).

Understanding the differences in crash rate and crash frequency is critical to implementing policy beneficial for safety in active transportation. Recent increases in pedestrian and bicycle fatalities in the United States have coincided with urbanization, and not surprisingly those fatality increases have been most prominent in urban areas, especially on major arterials (Cicchino & Hu, 2018). One study modeled pedestrian crash frequency and pedestrian crash risk separately, finding that walking mode share was positively associated with crash frequency but negatively associated with crash risk (Chen & Zhou 2016). If crash count is the variable of interest, the logical solution may be to discourage active transportation modes or further separate vehicles. If crash rate is the focus, more active transportation and traffic calming may be the answer. The interaction between pedestrian crash rate and crash frequency is often tied to the concept of “safety in numbers”, which is discussed further in later sections.

2.2.1.2. Speed and frequency of crashes involving bicyclists and pedestrians

There is a clear gap in the literature where vehicle speed and crash frequency ought to be addressed. Several studies incorporate speed limits into their analyses (Ma et al. 2010, Zahabi et al. 2011, Siddiqui et al. 2012, Chen 2015, Chen & Zhou 2016, Pirdavani et al. 2017), but few consider actual operating speed. One exception is Quddus (2008), in which average speed is back-calculated from congestion figures and included in the crash frequency model.

When included in area level bicycle and pedestrian crash studies, speed limit is generally a positive and significant explanatory variable in crash frequency (Chen 2015, Ma et al. 2010,

Siddiqui 2012). However, speed limit has a complex relationship with operating speed (Parker 1997) and in the American context is often tied to roadway functional class. We cannot conclude that including speed limits adequately addresses the effect of vehicle speed profiles in studies such as these.

When speed is considered at all, it is typically incorporated as a single speed benchmark that describes the overall speed level (as in the case of Quddus (2008) or studies using posted speed), not the nature of the speed distribution itself. Speed distribution is rarely addressed in pedestrian or bicycle crash literature.

2.2.1.3. Severity of bicycle and pedestrian crashes

In contrast to crash frequency studies, speed is commonly the primary explanatory variable in pedestrian and bicycle injury severity studies. Pedestrian crashes are more commonly studied than bicycle crashes. These studies generally model the probability of more severe crash outcomes in a pedestrian crash, given a pedestrian crash has taken place.

Kroyer et al. (2014) conducted a review of studies that model probability of pedestrian fatalities, finding that the most-cited studies use one of two data sets as the basis for study: one collected by Tefft (2013) and another by Rosen and Sander (2009). These studies generally show that the fatality risk curve is S-shaped; that is, risk of severe crash outcomes increase quickly with speeds when speeds are lower, reaches an inflection point, and slows at higher speeds. Tefft thought this inflection point occurred around 40 mph for fatality crashes (Figure 5).

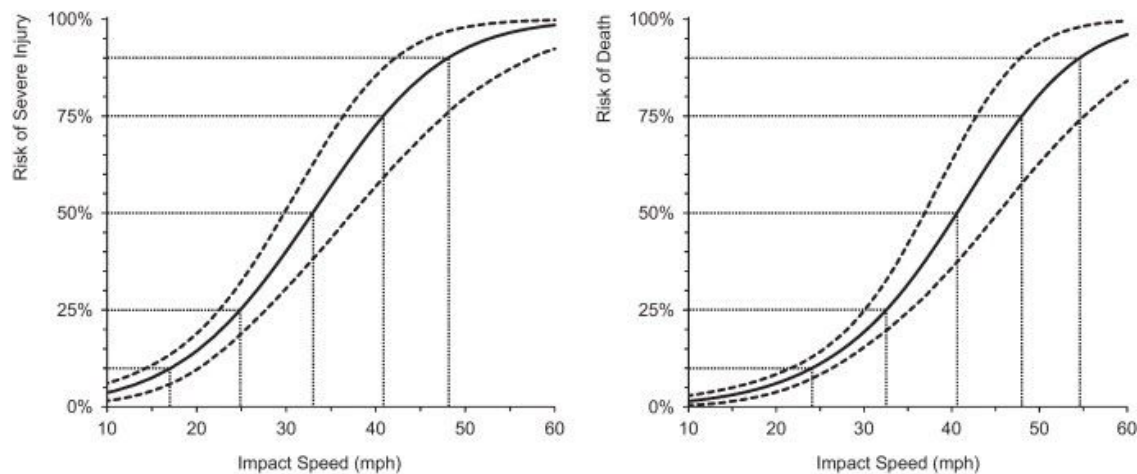


Figure 5: The S-shaped severe injury and fatality risk curves modeled by Tefft (2013)

Kroyer et al. note that the likelihood of a crash is not as practically important as total crash risk, or likelihood and exposure. Urban settings have both lower speeds and more pedestrian exposure, so they are likely to have greater pedestrian safety benefits to speed reduction and should be considered the primary target areas to reduce vehicle speeds.

Common data sources for studies of this kind are vehicle black box devices (Tefft 2013, Song et al. 2017), crash records themselves (Oikawa et al. 2016), and crashes linked to medical records (Tarko & Azam 2011). Each comes with its own challenges. Black box devices sometimes record vehicle speed (Song et al. 2017), but some require analysis of video at impact to calculate speed (Tefft 2013). Crash records place faith on the responding officer to accurately estimate the speed of the vehicle on impact, which has been shown to be inconsistent (Oikawa et al. 2016).

In studies linking medical records to crash outcomes, selection bias is an important consideration. Namely, crashes with a hospital record are likely to be more severe than the set of

all crashes, and this non-random sampling may introduce bias into the estimation. Tarko & Azam (2011) used medical and crash data to model the likelihood of severe injury and showed that bivariate probit models can be calibrated to account for this selection bias. Those models outperformed univariate probit models, which did not account for such bias.

In some cases, crash frequency and crash severity are modeled together. This is often achieved by separating crashes by severity and modeling crash frequency on each set of crashes separately. Researchers can then draw conclusions about which variables have greater or more significant impact on severe crash probability (Prato & Kaplan 2015, Noland et al. 2008). Intuitively, vehicle speed may have a greater positive relationship with severe bicycle and pedestrian crash probability than less severe crashes (Quddus 2008), but more research is needed.

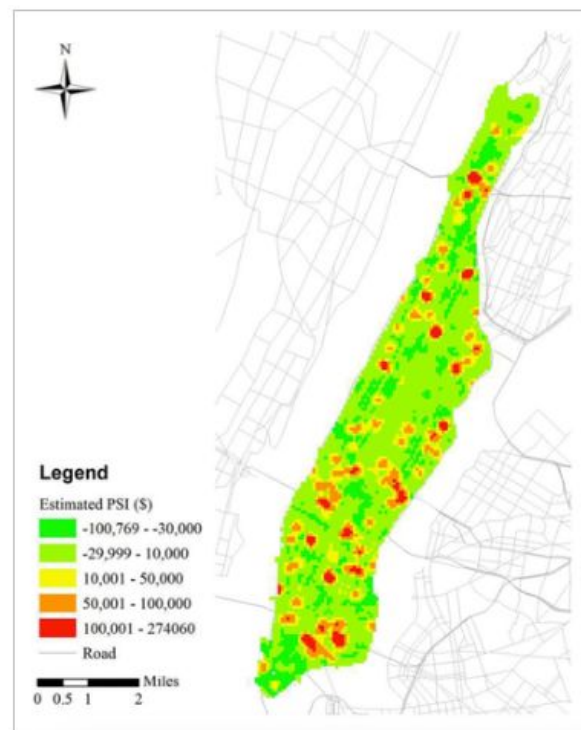


Figure 6: Area-level social crash cost in Manhattan, as shown in Xie et al. (2017)

Xie et al. (2017) took a unique approach to modeling crash probability weighted by severity. They converted crash severity to crash cost (the social cost incurred by an injury, debilitating injury, or fatality as reported by the National Safety Council), and modeled crash cost as the dependent variable in an independent study (Figure 6).

2.2.2. Crash count modeling techniques

Crash modeling techniques continue to develop today. Linear models have long been shown to be inadequate, since crash outcomes are discrete and at least zero. As a result, Poisson and Negative Binomial (NB) models have long been the standard for estimating crash frequency (Lord 2005). The Negative Binomial model is generally preferred, because the Poisson model makes a restrictive assumption that the dependent variable is not over-dispersed (its variance is greater than its mean). In the case of unlikely, random events such as crashes, the Negative Binomial model offers welcome freedom from that restriction (Lord 2005).

Crashes are relatively rare events, and depending on the unit of analysis, many observations may have zero crashes. Zero-Inflated Negative Binomial (ZINB) have become popular for modeling this kind of data. These models break out observations with zero crashes from observations with at least one crash, modeling them separately (Lord 2005). These models have shown improved fit compared to NB models in some cases, because aggregated crash data frequently comes with a preponderance of zeros. However, they come with an important caveat. ZINB models assume that 0 crash observations are “perfectly safe”, that their roadway characteristics created zero crash conditions, not random chance. For this reason, some researchers have considered ZINB models with some skepticism (Lord 2005).

More recently, Bayesian methods have grown more popular, and have become the new standard for crash prediction modeling (Saha et al. 2015). The key advantage of Bayesian methods is their ability to account for spatial correlation in crash occurrences. Thus, Bayesian methods can account for a significant portion of the variation found in area levels crash data (Jovanis 2006, Quddus 2008, Chen 2015). Several studies have compared naïve Negative Binomial models to similar models with Bayesian effects and found that the latter had significantly greater explanatory power in modeling crash frequency (Quddus 2008, Siddiqui et al. 2012). Importantly, Quddus (2008) found that when comparing naïve NB models to Bayesian, most variable coefficients were similar, except for the coefficient for average speed. This finding suggests that spatial random effects may be correlated with vehicle operating speed.

Functionally, accounting for spatial correlation in crashes requires creating a variable for each unit that captures the crash levels of adjacent units. At its simplest, the crash behavior of adjacent neighbors is accounted for, but a weighted nearest neighbor is more common (Aguero-Valverde & Jovanis 2008), Figure 7 shows an example of weighted nearest neighbor logic for nearby road links.

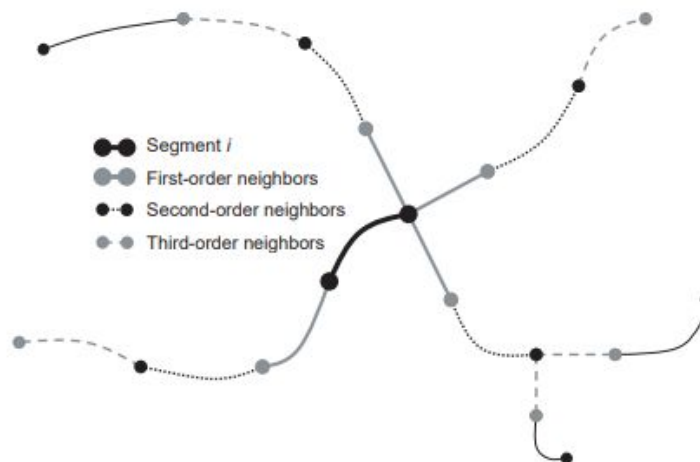


Figure 7: Weighted nearest neighbor on nearby road links (Aguero-Valverde & Jovanis 2008)

Tasic et al. (2017) propose the Generalized Additive Model (GAM) for modeling crashes to account for spatial correlation in crash data. They argue that the modeling process' key advantage is being simpler and less intensive than Bayesian counterparts.

2.2.3. Built environment factors and crash probability

2.2.3.1. Roadway infrastructure characteristics

Several characteristics of the roadway and traffic flow are generally included in crash frequency studies. Intuition and significant research indicate that the interaction among drivers and between drivers and their immediate environment greatly influence the chances of a crash. Several studies that evaluate these factors are discussed here.

Traffic volume, intersection density, and roadway functional class are often included as controlling variables in studies like these. Greater traffic volume leads to increased potential conflict, both among vehicles and between vehicles and active transportation users (Kaplan & Prato 2015). Some of the earliest speed and safety studies did not control for traffic volume, leading to erroneous results (Baruya 1998, Dumbaugh & Rae, 2009).

Intersection density is a measure of conflict with vehicles (Quddus 2008, Dumbaugh & Rae 2009) and is considered especially important in bicycle crash models (Strauss et al. 2015, Kaplan & Prato 2015). Vehicle access points (driveways and other curb cuts) have also been studied and found to be related to bicycle and pedestrian crash frequency (Vandenbulcke 2014).

Incorporating the effects of roadway functional class is especially important now that major arterials are seeing the bulk of increases in pedestrian injuries and fatalities (Cicchino & Hu 2018). Functional class can be studied in many different ways. Some area-level studies

include number of miles of arterial roadways as an explanatory variable (Dumbaugh & Rae 2009). Teschke et al. (2012) found that local streets carried about half the risk of a bicycle crash compared to higher functional classes.

Pedestrian and bicycle infrastructure are often studied to understand their degree of efficacy in improving bicycle and pedestrian safety. Pedestrian crossings, whether at intersections or midblock crossings, have been studied with mixed results (Ma et al. 2010, Tarko & Azam 2011). Many bicycle studies focus on the effect of different bicycle infrastructure types. Generally, these studies find that separated facilities show reductions in crash risk (Teschke et al 2012). Perhaps surprisingly, contraflow bicycle facilities on one-way roads have shown reduced crash risk (Vandenbulcke 2014). Some area-level studies identify crash hotspots for bicycles to identify priority bicycle facility locations (Strauss et al. 2015).

The width of the roadway, whether right-of-way width or number of lanes, may also be a factor in explaining active transportation crashes. This element of the roadway may affect the bicyclist or pedestrian's ability to judge a safe crossing or turn, as well as having an influence on vehicle speeds (Ma et al. 2010). Ukkusuri (2012) found that the number of lanes was positively associated with pedestrian crash frequency, and that right-of-way width was less effective at explaining crash frequency.

2.2.3.2. Land uses, land use types, and urban form

Several authors point out that infrastructure elements alone are not sufficient to design effectively for pedestrians and bicyclists, and that greater design principles like land use and urban design have a more significant impact (Cervero & Kockelman 1997, Dumbaugh & Rae

2009). In essence, the roadway does not exist in isolation, but interacts with the surrounding environment, and many of its outcome characteristics are a result of those surroundings (Cervero & Kockelman 1997).

Many researchers include surrounding land uses in their models. Considering types of land use (residential, commercial, industrial) is common, and intuitively these different uses capture details in pedestrian and bicycle travel patterns as well as the nature of conflicts with vehicles. Frequently, industrial and commercial uses are associated with increases in pedestrian crash frequency, and residential land uses are associated with relatively lower crash frequency (Ukkusuri 2012). Crash frequency for pedestrians tends to increase closer to the city center (Kaplan & Prato 2015).

The way that studies incorporate land uses vary widely, however. While many studies incorporate simply commercial, industrial, etc., Ma et al (2010) consider both land use type and intensity. They separate primarily residential land uses and scattered residential land uses categorically.

Some literature suggests that balance among different uses is a factor in pedestrian and bicycle crashes. Wang & Kockelman (2013) found that land use entropy, a measure of balance among residential, commercial, office and industrial land uses (versus a dominating single land use, for example), is negatively associated with pedestrian crash frequency. In this study, a more balanced distribution of uses improved pedestrian safety conditions. Chen (2015) found that greater land use mix has a positive impact on bicycle crash frequency.

The planning literature adds to this area of study by considering urban form as well as land use type. Cervero & Kockelman (1997) show that the “3 D’s” (density, design and

diversity) have a meaningful impact on pedestrian and automobile trip-taking, with population density being the greatest factor in promoting pedestrian activity and reducing auto activity.

Notably, Dumbaugh & Rae (2009) examine the design of commercial space and its interaction with pedestrian safety. They separate commercial development styles, identifying pedestrian scaled commercial areas (common in pre-World War II street corner development) with the more auto-oriented arterial commercial developments. The former focuses on smaller street-fronting development accessed from the sidewalk, while the latter is usually anchored by a “big box” store and often places parking between the street and the storefront. Their study found that the arterial style commercial development was associated with an increase in pedestrian crashes, while the pedestrian scaled commercial is linked to a reduction in pedestrian crashes. The authors point to differences in the nature of conflict between pedestrians and cars, as well as a road design that promotes high speed in the first case and lower speeds in the second.

2.2.3.3. Congestion

There is a complex relationship among congestion, vehicle speed, crash frequency, and crash severity. Intuitively, congestion and vehicle speeds are negatively correlated. Without accounting for congestion, it appears (as it has in many studies in the past) that higher speeds lead to fewer crashes. This is what explains the low end of Solomon’s famous U-shaped curve. In reality, it is not the lower speed that leads to more crashes, but the greater incidence of conflict caused by congestion (Baruya 1998).

Congestion also affects crash totals and crash severity differently. In a study estimating the probability of fatality and severe injury in pedestrian collisions in Hong Kong, Sze et al.

(2007) found that moderate and severe congestion reduced severity outcomes. The authors point to reduced speeds on impact, as outlined in sections above, as the mitigating factor. Noland & Quddus (2005) also found that peak hour conditions in London reduced fatal and severe crashes overall, but less severe crashes were more common at this time.

To make matters more complicated, there is some evidence that congestion mitigates bicycle and pedestrian crashes in another way—by discouraging active transportation trips. Noland et al. (2008) performed a bicycle crash before-after analysis of London’s congestion charge, and found that counts of severe bicycle injuries actually increased. The authors estimate that not just higher traffic speeds, but greater bicycle traffic caused greater exposure to injury crashes for bicyclists. Perhaps drivers priced out by the congestion tax chose to take bicycle trips, or perhaps more people chose to travel by bicycle with fewer cars to compete with.

More research is needed to understand the complex relationship between congestion and bicycle and pedestrian crashes. However, one fact is clear; one must account for congestion to adequately assess the impact of speed itself on area-level crashes.

2.2.3.4. Safety in numbers: bicycle and pedestrian activity and crash rate

An important line of research explores how bicycle and pedestrian activity and crash rates are related. The phenomenon has been coined “safety in numbers”, and suggests that promoting active transportation trips is itself a safety measure. If more bicyclists and pedestrians are out on the streets, the thinking goes, then drivers will grow more aware of them and modify their driving behavior to account for their safety. Geyer et al. (2006) showed that pedestrian

activity increases at a greater rate than pedestrian crashes, so crash rate decreases with greater pedestrian activity. The phenomenon is common in bicycle literature as well (Tasic et al. 2017).

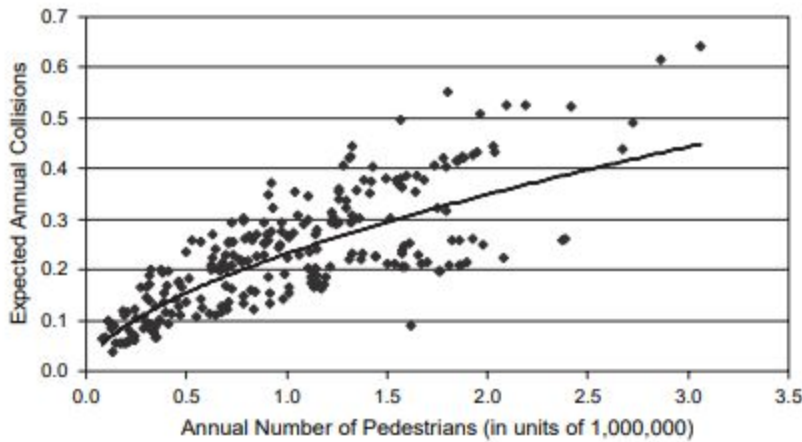


Figure 8: Sub-linear growth rate between pedestrian activity and pedestrian crash counts at 247 intersections in Oakland, CA (Geyer et al 2006).

A key concern with this line of research is the causal link between activity and safety. Specifically, underlying infrastructure design could affect both activity and safety. For instance, a roadway designed with fewer lanes and for lower vehicle speed might reduce pedestrian crashes on its own and also encourage pedestrian activity through a perception of a safer, more pleasant urban environment. While this question has yet to be completely resolved, active transportation trips have been clearly identified as an important factor in crash frequency.

2.2.3.5. Public assets: transit, schools, and parking

Various public assets along the roadway are considered in the crash frequency literature because they tend to act as indicators of special conditions like concentration of active transportation users, complex environments, or concentration of children or the elderly. Most

commonly, transit assets and schools are considered, but parks and parking space are also explored.

The presence of schools and parks is often studied in conjunction with pedestrian crashes (Xin et al. 2017, Zahabi et al. 2011). Pedestrians and bicyclists, especially children and teens, are attracted to these environments, so intuitively conflicts with vehicles are more common near their borders. Xin et al. (2017) found that proximity to parks and schools were significant factors in risk of a pedestrian injury in a collision.

The presence of on-street parking is frequently studied as a risk factor in bicycle and pedestrian crashes. In one bicycle study, on-street parking doubled the risk of collision on major roads (Teschke et al. 2012). In a simulated driving study, driver speed behavior was studied in relation to the presence of on-street parking. With the presence of on-street parking--and pedestrians accessing those vehicles--drivers generally lowered their speed. However, that compensation was not large enough to safely avoid conflicts that arose. Overall, risk of a crash increased in these environments despite driver speed reduction. The authors of this study concluded that drivers do not sufficiently compensate for the risks of complex urban environments (Edquist et al. 2012).

Transit stops and transit access are commonly included in macroscopic models of pedestrian or bicycle crash counts (Ukkusuri et al. 2012, Wang & Kockelman 2013, Chen & Zhou 2016, Xin et al. 2017, Tasic et al. 2017). These variables are frequently used as a proxy for pedestrian—or, less commonly, bicycle—activity. Going further, researchers consider transit access to be a potential cause of conflict between active road users and vehicles. Transit stops become crossing points for people attempting to make the bus or train. However, this effect

independent of pedestrian activity is not clear in the literature. Wang & Kockelman (2013) found that after controlling for pedestrian activity, bus stop density was not significantly related to pedestrian crashes.

2.2.4. Socioeconomic and sociodemographic characteristics

Traditionally, most factors included in crash studies have been elements of the built environment or descriptive components of how the transportation system is being used (e.g. vehicle miles traveled, mode share, etc.). Demographic characteristics have become more common in area level crash studies, however, because they are readily available when the unit of analysis is a political boundary (e.g. census tract or Traffic Analysis Zone), and because they have recently been shown to be significant (Dumbaugh & Rae 2009, Ukkusuri 2012, Siddiqui 2012, Pirdavani et al, 2017, Xin et al. 2017).

One key demographic factor is income, which at the area level is nearly always negatively associated with bicycle and pedestrian crash frequency (Pridavani et al. 2017). This factor is correlated with car ownership in the U.S. and factors into mode choice.

Gender is also considered an important factor; male pedestrians and bicyclists are often more risk-loving and are more often victims in a pedestrian or bicycle crash (Pirdavani 2017).

Finally, race has proven to be significant in several area level crash studies. This factor is complex due to its entanglement with income and community design, along with cultural factors. Traffic Analysis Zones (TAZs) and census tracts tend towards internal racial/ethnic homogeneity (Xie, et al. 2017), a fact largely due to practices of segregation and redlining throughout American history. For example, Xin et al. (2017) found that speed limits are lower in black

neighborhoods in Florida because these neighborhoods tend to be older and more centrally urban. This finding confounds the speed and safety relationship at the area level.

Socioeconomic and sociodemographic characteristics interact with important components of the speed and safety relationship, especially when focusing on bicycle and pedestrian crashes (Pirdavani 2017). Accounting for these factors in area-level crash studies may reduce error when modeling the relationship between speed and safety.

2.3. Summary

A review of the literature revealed that more research is needed to understand the impact of vehicle speeds on bicycle and pedestrian safety. Vehicle speeds are well-studied in relation to frequency of crashes generally and severity of collisions with pedestrians. However, most studies of bicycle or pedestrian crash frequency have not adequately incorporated vehicle speed. The literature has identified other important explanatory factors in bicycle and pedestrian crash frequency. Namely, traffic volume, congestion, roadway characteristics, population density and demographic characteristics, bicycle/pedestrian activity, and land use characteristics have been shown to be critical factors. Incorporating these factors is necessary for the success of an area-level crash study.

Chapter 3

Data

The data used in this analysis chiefly comes from two sources. Crash data comes from the Georgia Department of Transportation (GDOT), which processed and provided a spatial crash data set for the years 2013-2017. Probe vehicle speed data was accessed from the National Performance Measures Research Dataset (NPMRDS). This data set is provided and maintained by INRIX, Inc. and accessed via license with GDOT. INRIX began providing speed data for NPMRDS in 2017. Thus, 2017 was the only year with complete data for this analysis. Other supplemental data sources were included to complete the analysis.

In what follows, the data sources are described in detail, then the methods used to prepare the final data set are outlined. Figure 9 shows a flowchart of steps in data preparation and the inputs needed to complete those steps. All data preparation was completed in R version 3.6.1 and ArcMap®.

3.1. Data sources

3.1.1. Crash data

The Georgia Electronic Accident Reporting System (GEARS) crash database is an exhaustive collection of crash records submitted in the state of Georgia. The records are completed by police officers responding to the crash, and the data set is compiled and maintained by GDOT. The database is web-hosted, so an end user can query the database to access crash records by date, time, jurisdiction, vehicle type, and other information listed on the crash record.

GDOT also creates and maintains a spatial dataset of crashes, which has been pre-processed for spatial analysis. The spatial dataset was the primary crash resource used in this analysis, but direct pulls from GEARS were used to backfill and verify the data set. The backfill process is described in the data preparation section.

3.1.2. Speed and roadway data

Roadways and vehicle speeds come from the NPMRDS probe vehicle speed data set. These data were accessed via the Regional Integrated Transportation Information System (RITIS) platform. Speeds are available as a 5-minute harmonic average for each monitored road segment in the Traffic Message Channel (TMC) system. This TMC system breaks the roadway into unique segments designed for effective delivery of traffic and travel information (NPMRDS). End-users can specify the road segments of interest and the time boundaries to receive speed data, and request a download. The data is delivered in a comma-separated ASCII text file, where one row represents vehicle average speed for a five minute interval in a TMC. Each row contains a timestamp, a TMC code, and an average speed. The RITIS platform also provides a spatial data set of all TMCs for which speed data is collected--a global TMC shapefile--for each U.S. state. This data contains roadway attributes such as lane count, AADT, functional class, etc. TMCs are represented as lines.

TMCs are defined as either internal or external. External TMCs are stretches of roadway between major intersections or junctions, while internal TMCs are the short segments that intersect or fly over intersecting roads. Internal TMCs mark a transition point on major roadways

and freeways around which large portions of traffic may enter or exit the roadway. See Figure 9 for an illustration.

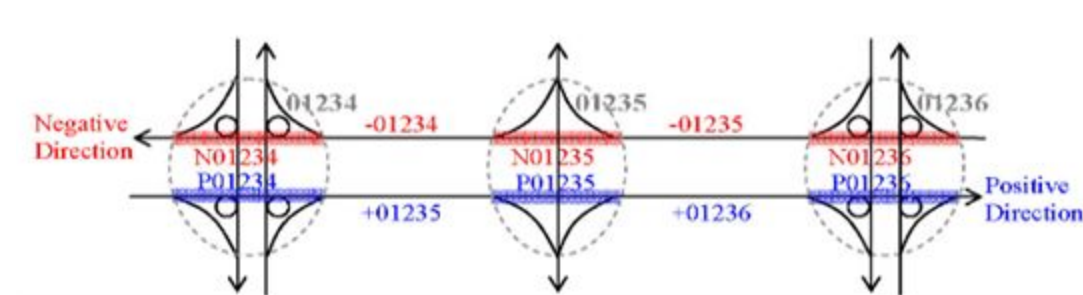


Figure 9: Example of Internal (Red/Blue), External (Black) TMCs (NPMRDS Analytics)

It is worth noting that external TMCs can be adjacent and do not have to be separated by internal TMCs. Admittedly, the reasoning behind breaks between adjacent TMCs is not always evident.

3.1.3. Speed limit data

Speed limit data was provided by GDOT, in the form of a spatial data file from Google Earth® (.kmz). The data set contains a network of Georgia roadways represented as lines, and each line contains the road segment's speed limit attribute.

3.1.4. Climate data

Climate factors are important for crash prediction, since they vary in terms of rainfall and fog patterns, which affect visibility and pavement conditions (Taylor 2000). Data on ecoregions in the continental United States is publicly available from the EPA. Land within these ecoregions share important climatological and ecosystem characteristics (EPA 2018). The climatological

characteristics themselves are not included in the final data set; rather, ecoregion names serve as indicators that TMCs experience similar weather patterns, if in the same ecoregion, or different weather patterns, if in different ecoregions. Ecoregions are represented as polygons.

3.2. Data preparation

Figure 10 below shows a diagram of the steps in data preparation and the data inputs used to complete those steps. In this section, each of these processes is described in detail.

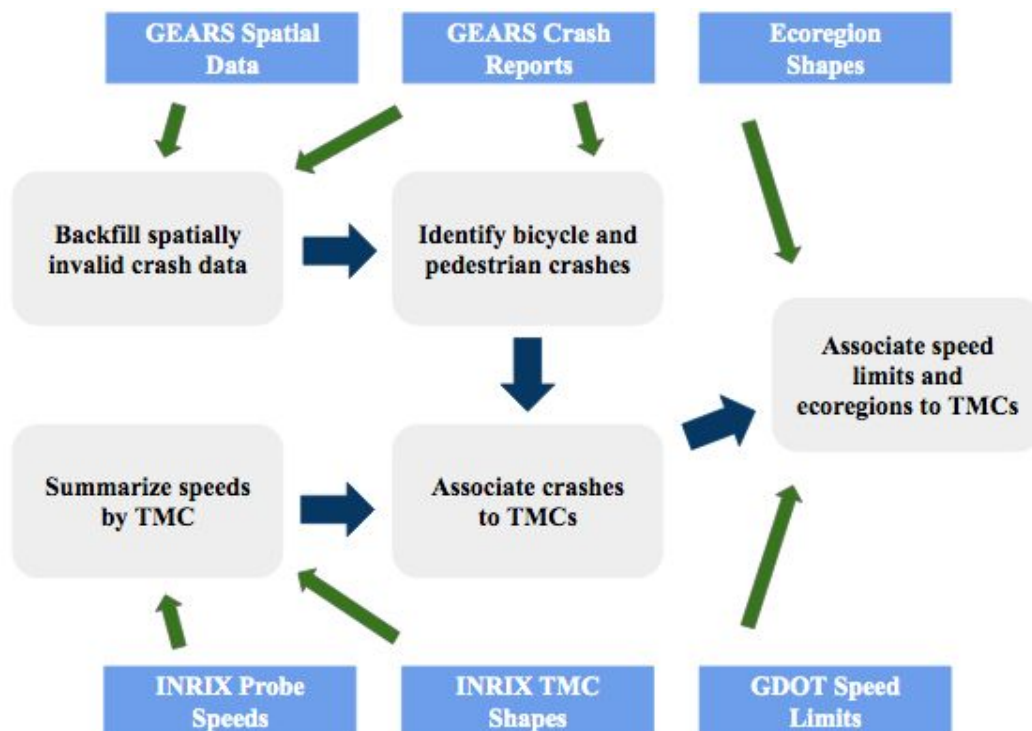


Figure 10: Steps in data preparation and their necessary inputs

3.2.1. Crashes: backfilling invalid spatial data

The GDOT spatial crash data set contains 412,035 crashes in the year 2017. Within the full crashes spatial data set, about 5% (24,000) of crashes were identified as having invalid

spatial data. That is, their data points fall well outside Georgia's boundaries, or they do not map successfully. Bicycle crashes were disproportionately affected by this issue: almost 100% of 2017 bicycle crashes did not map successfully. To address this issue, the latitude and longitude from the GEARS online database were used to backfill spatially invalid data.

To backfill this data, all 2017 crash records were pulled from the GEARS online database, which contains text latitude/longitude (lat/lng) attributes. These records were joined to the invalid spatial data by crash ID. Roughly 19,000 of the 24,000 invalid points successfully joined with their GEARS record. At this point, the lat/lng text from the GEARS record was read as spatial data and verified spatially. Roughly 11,000 of the 19,000 points were successfully mapped inside the state of Georgia. To confirm the accuracy of the new data, 100 of these 11,000 points were randomly selected for manual verification: the location described in the crash record itself was compared to the map location of the newly backfilled spatial information. All 100 points matched correctly, so the 11,000 successfully-backfilled crashes were used in the final data set. At this point, 399,147 crashes were included.

3.2.2. Identifying bicycle and pedestrian crashes

Since bicycle and pedestrian crash identifiers were not present in the GEARS spatial dataset, flags were created in the spatial data set for both bicycle and/or pedestrian crashes. To create these flags, the GEARS crash database was queried to identify crash IDs that represent bicycle or pedestrian crashes, and those crash IDs were flagged in the spatial dataset.

For pedestrian crashes, if the reporting officer checked a box on the crash report indicating a pedestrian was involved, the crash record was flagged pedestrian and was included.

4,406 pedestrian crash records from 2017 were identified in GEARS using this method. Manual verification of ten randomly-selected crash records revealed that records with the pedestrian indication consistently described crash events involving a pedestrian.

For bicycle crashes, the report was included if the vehicle type of any vehicle involved in the crash event was listed as bicycle or pedacycle. The bike indicator checkbox was not used, since it was implemented during the year 2017, so crashes could not be identified with the bicycle indicator flag before a certain date in late 2017. Furthermore, once the bike indicator checkbox was added to the form, it was used inconsistently: ten randomly selected crash records with the bike indicator checkbox were manually reviewed, and eight did not involve a bicycle. In contrast, nine of ten randomly selected crash records identified by vehicle type involved a bicycle. In all, 731 bicycle crash records from 2017 were identified in GEARS by vehicle type and included in the analysis.

Completely and accurately identifying crashes involving a pedestrian or bicyclist was not possible due to the many differences in the way the crash reports were filled out. In addition, the methods above would not identify false negatives, or crashes involving a bicycle/pedestrian but not flagged as such. However, we believe the methods described above to be relatively accurate.

Once identified from the GEARS database, crash records involving a pedestrian or bicyclist were flagged in the spatial data set by matching the crash ID. Overall, 3,675 of the 4,406 pedestrian crashes and 691 of the 731 bicycle crashes were successfully flagged in the spatial data set. The bicycle and pedestrian crash count in the spatial data set was 4,239.

Figure 11 shows the distribution of bicycle and pedestrian crashes across the state of Georgia. Crashes are not evenly distributed in space; they are clustered in population centers,

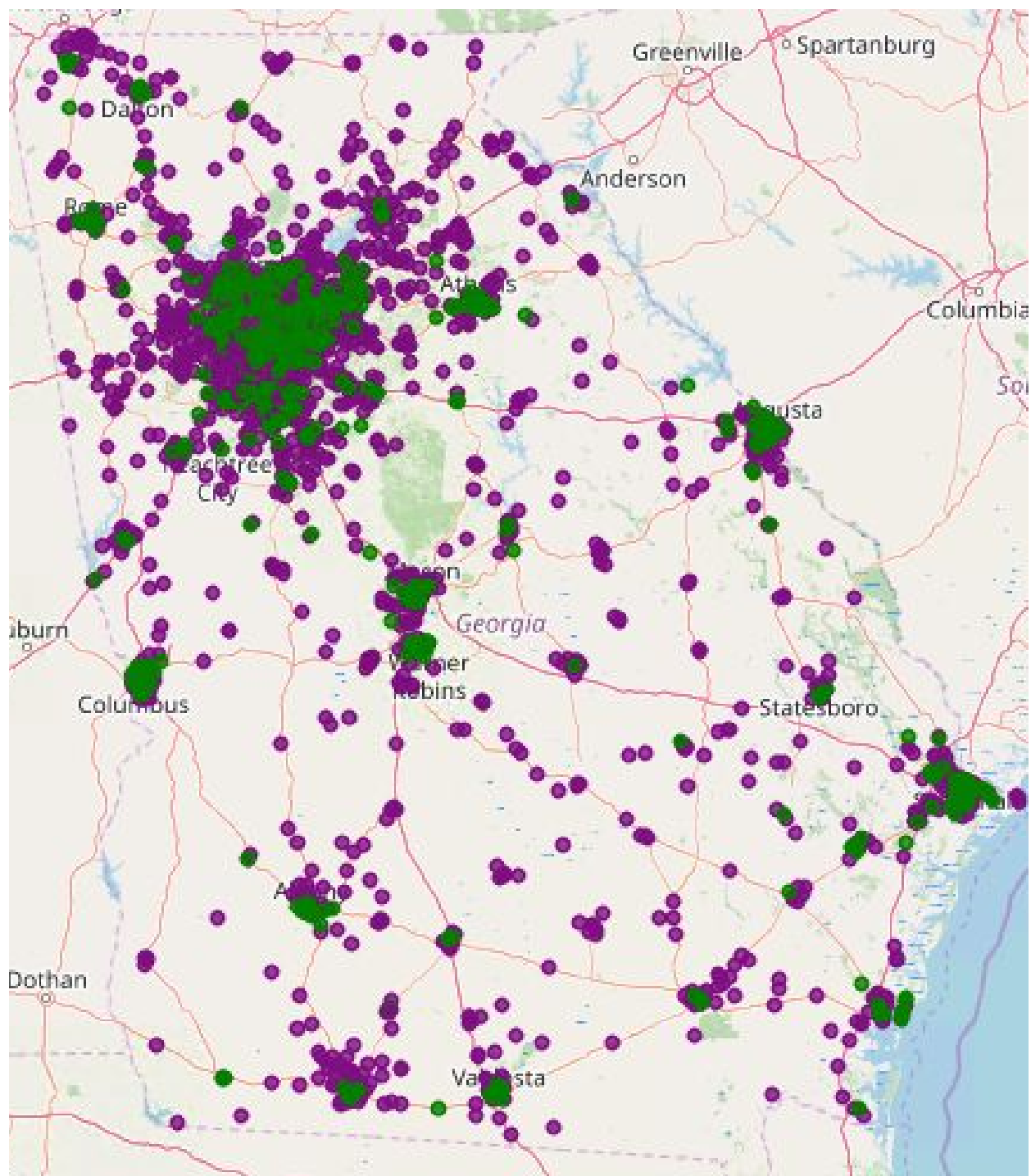


Figure 11: Bicycle (green) and pedestrian (purple) crashes cluster with population in Georgia

where more vehicles, bicyclists, and pedestrians are present. Georgia has some dense population clusters (e.g. Atlanta, Savannah, Columbus, Macon) alongside large rural areas. Figure 11 highlights the importance of controlling for population and demographic characteristics when evaluating the effects of roadway and speed characteristics on crash frequency. Otherwise, large variation in population density can confound the speed-safety relationship.

3.2.3. Processing TMCs and speeds

To create a spatial data set of TMC road links containing 2017 speed summaries, speed data was summarized by TMC and combined with spatial data. Figure 13 shows the included TMCs and their relative 85th percentile speeds. TMCs were considered for inclusion if probe vehicle speeds were collected in 2017. The general cutoff for probe vehicle speed collection was 13,000 AADT or greater (based on agreements between GDOT and RITIS), but some TMCs with lower AADT than the cutoff had probe vehicle speed data due to operational considerations.

Table 1: Included TMCs by Functional Class

Functional Class	Count
3. Major Arterial	6,318
4. Minor Arterial	640
5. Major Collector	74
6. Minor Collector	0
7. Local	18
Total	7,050

First, pertinent TMCs were selected from the statewide TMC shapefile. TMCs with a functional class of interstate highway (1), other freeway (2), or ramp (8) were excluded, since they are not appropriate for bicycle and pedestrian use. The 83 TMCs with fewer than 1,000 speed observations in 2017 were also excluded.



Figure 12: Non-primary TMCs overlap with primary TMCs (NPMRDS Analytics)

Finally, only primary TMCs were included. TMCs are identified in the NPMRDS as primary or not primary, and those not identified as primary are spatially redundant with primary TMCs (NPMRDS). Figure 12 shows an illustration of the overlap scenario between primary and non-primary TMCs. After data preparation, 7,050 TMCs were included in this analysis (Table 1).

Importantly, even primary TMCs may overlap. The most common case for TMCs to overlap is when two lines representing the same road segment--but opposite directions of travel--have similar or identical spatial information (i.e. represented by identical line segments). All TMCs are directional, and most direction pairs are represented side-by-side in space (e.g. a highway divided by a median), but many TMCs pairs, especially those that represent city streets, are spatially identical. Another, less common case of overlap is the crossing of two lines representing intersecting roadways. These are both considered valid cases, since they accurately represent the topology of the roadway, so their overlap was accounted for in later steps.

The raw speed observations were then summarized and attributed to included TMCs. All of the 5-minute speed observations for each TMC were summarized into speed percentiles

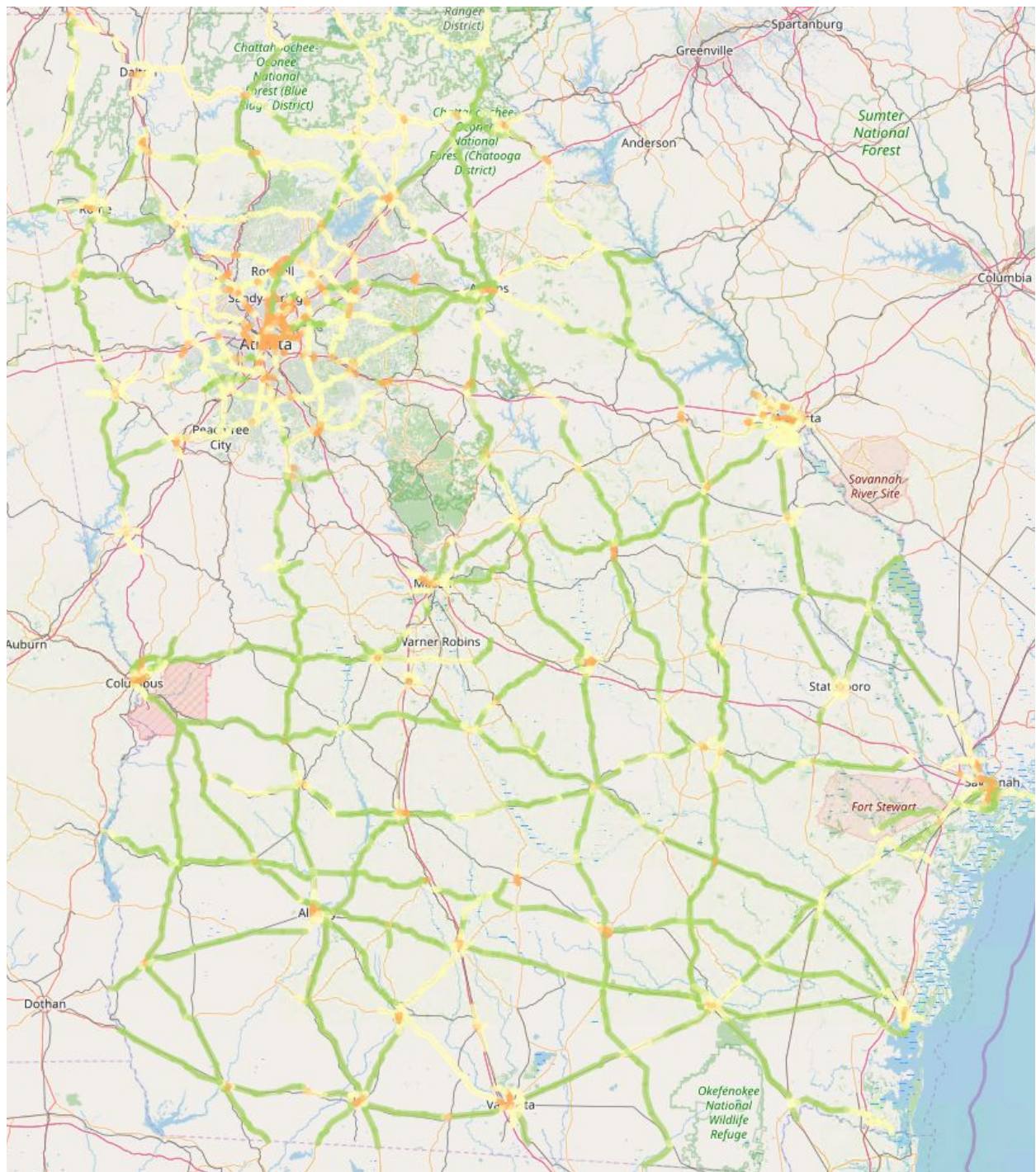


Figure 13: Included arterial roadways in Georgia, symbolized by 85th percentile speed. Lowest values (orange) and low values (yellow) are concentrated in urban centers, and higher values (green) are more common in rural areas.

(5th, 10th, 15th, etc.) for the year 2017. These percentile speeds were then attributed to the spatial TMC data set by matching TMC codes. Figure 13 shows the spatial TMC data, symbolized by 85th percentile speeds. The fastest speeds (green) are mostly found in rural areas, while the slower (yellow) and slowest (orange) speeds are more often found in urbanized areas. Figures 11 and 13 together illustrate that population density can confound the speed-safety relationship.

Measures of speed distribution were also computed, since this study focuses on the distribution of speed on a roadway as a causal factor in bicycle and pedestrian crashes. Following Ederer et al. (2020), the difference in 85th and 50th percentile speed (hereafter referred to as the high speed difference) was calculated to represent the upper distribution of speeds, and the difference between 50th and 15th percentile (the low speed difference) was calculated to represent the lower distribution of speeds.

3.2.4. Conflating speed limits to TMCs

As mentioned above, GDOT provided posted speed limit (PSL) data in the form of polylines for the Georgia roadway network. Since no attributes uniquely identified roadways in both data sets, and the two data sets overlap incompletely in space, a spatial association was required to attribute speed limit data to TMCs. This process is often referred to as spatial conflation. The spatial components were completed using ESRI ArcGIS® software, and the data management components were completed in R version 3.6.1.

The first step in the conflation process is to align the vertices of the two sets of polylines in space. A frequent challenge with line-to-line conflation is the imperfect overlap of two lines in

space, even though the lines are meant to represent the same physical object. ESRI's Integrate tool was used to align vertices of the TMC data and the speed limit data within a certain tolerance distance.

Choosing an appropriate tolerance is an art; too small, and vertices meant to represent the same object will not align, but too large, and vertices that represent different places will erroneously align. Thus, the smallest acceptable tolerance should be chosen. Table 2 shows conflation results at different tolerance levels. As the table shows, the difference in number of TMCs associated with a speed limit changes marginally between 35ft and 50ft. For this reason, 35ft was chosen.

Table 2: Results of posted speed limit conflation by tolerance distance

Value	25 ft tolerance	35 ft tolerance (Selected)	50 ft tolerance
TMCs with PSL (% of total)	5,722 (81.1%)	6,076 (86.1%)	6,123 (86.7%)
TMC miles with PSL (%)	10,021 (91.1%)	10,523 (95.9%)	10,544 (96.1%)
Bike/Ped Crashes with PSL (%)	822 (78.2%)	850 (80.9%)	847 (80.6%)

From here, the overlap between TMCs and speed limit polylines could be evaluated. It should be noted that because they were created for different purposes, road segments in the two data sets have different start and end points. ESRI's Intersect tool was used to divide unique pairs of overlapping TMC/speed limit polylines into unique observations. The attributes of both data sets were combined in the resulting data. Figure 14 shows an illustration of this process.

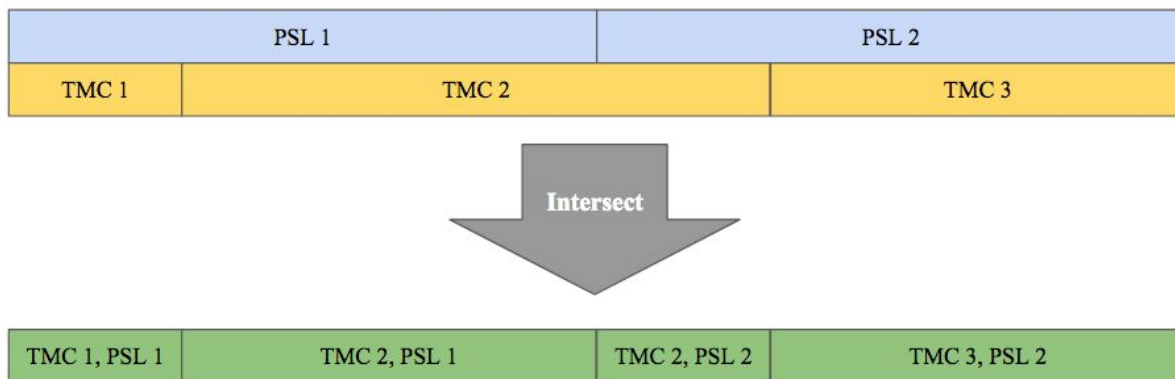


Figure 14: Illustration of Intersect process and resulting shapes and data attribution. Using the assumptions of this analysis, TMC 2 would be assigned PSL 1.

As shown in Figure 14, it was possible for multiple geometries in the output data set to have the same TMC ID. To get to a one-to-one relationship between TMCs and associated speed limit, the intersecting geometries were evaluated by TMC ID. The assumption was made that the intersect geometry with the greatest length was most representative of the whole TMC's posted speed. So, for each TMC ID, the posted speed limit of the geometry with the greatest length was selected.

After conflation, 6,076 of 7,050 included TMCs (86%) were given a speed limit attribute, but 96% of total TMC length was matched with a speed limit (Table 2). These figures, and visual confirmation, indicate that conflation with PSL data was more successful in rural areas than urban areas, where longer TMCs and less bicycle and pedestrian activity characterize the roadways. The primary reason for the incomplete conflation is lack of speed limit data in some areas of the network, but attempts to integrate geometries which were outside the tolerance range is another possible source of error. All TMCs that did not match to a speed limit through this process were manually assigned a speed limit via Google Street View.

3.2.5. Conflating crashes to TMCs

Since the unit of analysis in this study is the TMC, crashes were associated with TMCs and summarized to describe the number of crashes on each TMC in 2017. Location in space was the primary characteristic used to associate crashes with TMCs, but vehicle direction attributes of both crashes and TMCs were also evaluated to complete the process. This spatial conflation was completed entirely in R.

First, crash points less than 50 feet from any TMC line were considered to have taken place on a TMC and were included. Crashes within 50 feet but identified as taking place on an interstate (often “flying over” a non-interstate TMC) were excluded. 1,049 crashes (142 bicycle and 907 pedestrian crashes) were ultimately included.

At this point, crashes were associated with the TMC with the least distance between them. However, there were 228 cases where a single crash was associated with multiple TMCs by minimum distance. Most commonly, a crash point was located near two overlapping TMCs with opposite directions of travel, but in some cases a crash was located at the intersection of two TMCs. To break ties among these crash-TMC pairs, the crash vehicles’ direction of travel were compared to the TMCs’ direction of traffic flow, and the TMC-crash pair with matching vehicle direction was kept. The direction of vehicle one was considered first, then vehicle two. With this method, 124 of the 228 crash records associated with more than one TMC were resolved. If there was no match based on vehicle direction, the TMC was chosen randomly. The remaining 104 records were resolved by random choice. After this process, each included crash was associated with exactly one TMC.

3.2.6. Conflating ecoregions to TMCs

There are five ecoregions in the state of Georgia, and all of Georgia falls into exactly one of these ecoregions. TMCs were assigned ecoregions through a spatial join process. First, the center point, or centroid, of each TMC was found. Then, a TMC was assigned the ecoregion if its centroid falls within that ecoregion. Table 3 shows the count of TMCs by Georgia ecoregion.

Table 3: Included TMCs by Ecoregion

Ecoregion	Count
Blue Ridge	166
Piedmont	3,224
Ridge and Valley	439
Southeastern Plains	2,424
Southern Coastal Plains	797
Total	7,050

With ecoregions included, the data set used for this analysis was assembled. Table 4 shows the TMC-level summary of each variable of interest. Crashes are only present among a few TMCs: 756 of 7,050 TMCs were associated with at least one crash. Other road characteristics vary widely among TMCs. TMC length varies by several orders of magnitude, from a few thousandths of a mile to over twelve miles. AADT ranges from under 1,000 vehicles daily to over 100,000, and speed percentiles range by 60-70 mph. A broad variety of road types and contexts is present in the data set.

Table 4: Summary statistics for each TMC

	Mean	Median	Min	Max	Standard Deviation
TMC length (miles)	1.556	0.711	0.004	12.227	2.128
Crashes	0.149	0	0	8	0.511
Injuries	0.116	0	0	11	0.489
Fatalities	0.011	0	0	3	0.109
Speed (miles per hour)					
Posted Speed Limit	48.00	45	20	70	9.38
85th percentile	48.81	48	17	87	12.01
Median	40.13	39	9	68	14.10
15th percentile	29.99	27	4	64	15.54
High speed difference (85th – Median)	8.68	8	2	45	3.35
Low speed difference (Median – 15th)	10.13	10	2	36	3.77
AADT (vehicles per day)	17,800	14,641	531	119,000	13,604
Number of Through Lanes	3.59	4	1	8	1.13

Chapter 4

Methodology

This study seeks to estimate the impact of speeds, particularly speed distributions, on bicycle and pedestrian crashes and injuries. This study builds upon the methodology of Ederer et al. (2020), which found that differences in the speed distribution are a robust estimator of crash frequency in crashes on Georgia State route 6. Following Ederer et al., the study compares the results of a model in which the speed variables are absolute percentiles (15th, 50th, 85th) within the distribution to a speed difference model. In the speed difference model, the key independent variables are the high difference (between 85th percentile and median speeds), and the low difference (between median speeds and 15th percentile speeds) in the speed distribution. The comparison of these two models shows the effects of changes in speeds within a roadway's speed distribution versus the speed distribution itself.

4.1. Model specification

A Negative Binomial model is used. This model is the current standard for assessing the key factors in crashes because it assumes the dependent variable is a non-negative integer. Additionally, unlike the Poisson model, there is no restricting assumption on the dispersion of the dependent variable in the Negative Binomial model. Specifically, the Poisson model assumes that the dependent variable is not overdispersed, that its mean is equal to its variance.

The data included in this model is characterized by a preponderance of zeroes: only 756 of 7,050 TMCs were associated with at least one bicycle or pedestrian crash in 2017. The reason for this is understandable; bicycle and pedestrian crashes are naturally clustered in high

population areas (see Figure 11). The state of Georgia has significant rural land area characterized by very large distances between origins and destinations. In these areas, bicycle and pedestrian travel is often impractical, and total population density is low. It stands to reason that exposure to bicycle and pedestrian crashes is also very low. However, clusters of population density in Atlanta, Columbus, Savannah, Augusta, and Macon may have much greater exposure, despite having lower total land area.

The Zero-Inflated Negative Binomial model was considered to account for the preponderance of zeros. When applied to crashes, this model allows for a different specification for observations with zero crashes (perfectly safe road links) and observations with nonzero crashes (unsafe links). However, a key assumption for perfectly safe links is that either risk or exposure is effectively zero; that is, a bicycle or pedestrian crash could not occur due to the characteristics of the roadway (Lord 2005). This assumption is not realistic in this context; it is more likely that most links with zero crashes in 2017 had low (but non-zero) exposure and reasonable risk, and a crash event happened to not occur. As recommended in Lord (2005), the Negative Binomial model was used instead.

A key drawback of the Negative Binomial model in this context is its inability to explicitly account for the spatial nature of the data. Figure 11 shows that the location of the TMC is a critical factor as to how many bicycle and pedestrian crashes are concentrated on that TMC. Negative Binomial models can include attributes that characterize its location in space, like surrounding population density, AADT, and regional fixed effects, but they cannot account for unnamed, location-specific variables. This fact can be regarded as a weakness of this approach.

4.2. Other included covariates

Aside from speeds, the models tested in this analysis included controlling variables; the ultimate goal was a parsimonious model that addresses major crash factors at the network level. These variables were chosen based on the theoretical underpinnings of crash causation, practices within the literature, significance and magnitude of coefficients, and impact on measures of fit.

AADT is included as a controlling variable. AADT is log-transformed to reduce the large variation in their values across road segments; the difference between minimum and maximum value is several orders of magnitude (Table 4). Additionally, when these values are very large, small changes in their value may not be meaningful for crash prediction, but when the values are small, small changes may be more significant (Kaplan & Prato, 2015). The log transformation effectively addressed both of these concerns.

To control for the effects of the length of the roadway, TMC centerline mileage is included as a controlling variable. TMC length determines the time and distance a bicycle or pedestrian spends traversing the TMC, which indicates level of exposure. The length of the TMC is an extensive variable; that is, it defines one of the physical limits of the system. In this case, the sum of all TMC length describes the roadway length of the arterial network, and its relationship with crashes is constrained to be linear. Since it is not appropriate to perform a nonlinear transformation on extensive variables, TMC length is included directly.

In addition to TMC length, a boolean variable for short TMCs (defined as less than one half mile) is included in the models. A large portion of these TMCs are defined as internal TMCs (Figure 9). In general, short TMCs may have qualitatively different characteristics from external

TMCs aside from simply being shorter in length. This concept is developed further in the discussion section below.

Other key controlling variables are the posted speed limit, number of through lanes, and roadway functional class. These variables describe the road's context, traffic volume, and size, which can all have an impact on crash likelihood, and area level bicycle and pedestrian crash studies commonly include these variables in their analyses.

Number of through lanes in each direction is incorporated as a categorical variable, not a numeric variable. Intuitively, the difference between a one-lane road and a two-lane road, for example, is a categorical difference in the layout of the road, not a matter of degree. For bicyclists and pedestrians, the number of through lanes may change the perception of danger or discomfort in many different ways: pedestrians crossing multiple lanes or a bicyclist turning left may be qualitatively different depending on lane count. Bicycle and pedestrian activity may be affected as well.

Dummy variables for TMC's metropolitan area are also included. These variables begin to account for fixed effects specific to the region in which the crash takes place, like population density, local decisions in urban design and transportation planning, and cultural effects. While there is significant heterogeneity in these factors within a particular metropolitan area, accounting for each of these factors with greater granularity for the full state would require significant data collection and preparation effort.

Ultimately, climate data from ecoregions was not included in the results, since their inclusion proved to be inconsequential to the model results.

Table 5 shows correlations among numeric covariates included in the models. Correlation coefficients with a magnitude greater than 0.5 are highlighted.

Importantly, crashes and injuries are highly correlated (0.82). This result is consistent with the bicycle and pedestrian crash literature, and it reflects an important fact in studying bicycle and pedestrian crash frequency--crashes between a vehicle and a vulnerable road user almost always result in an injury or a fatality for the vulnerable road user. Reducing these types of crashes is nearly synonymous with reducing injuries and fatalities on the roadway.

The 15th, 50th, and 85th percentile speeds are each highly correlated (0.93 to 0.98). This result suggests that the speed distribution often shifts together when comparing across TMCs, and within a given TMC, a speed percentile is more indicative of the TMC's position along the speed distribution rather than the distribution itself. High correlations also suggest multicollinearity when many speed percentiles are included together in a regression model.

Percentile speeds are negatively correlated with AADT, likely indicating that increased volume generally coincides with congestion and lower speed. Length of TMC is also positively correlated with speed, which may suggest that longer TMCs indicate a simpler road context with longer spans between major intersections or landmarks which would warrant a new TMC.

Posted speed is highly correlated with 15th, 50th and 85th percentile speed. This result is not surprising, since posted speed is often set by the 85th percentile speed (NTSB 2017).

Table 5 also shows low correlation between crashes and high or low speed difference, emphasizing the need to control for other key TMC characteristics that may confound the speed and safety relationship. Overall, high speed difference has a greater magnitude of correlation with many key variables (crashes, injuries, AADT, TMC length) than low speed difference.

Table 5: Correlation matrix for included TMC variables, with values larger than 0.5 in yellow

	Crashes	Injuries	Fatalities	15th Speed	Median Speed	85th Speed	85th - 50th	50th - 15th	AADT	Length (mi)	Func Class	Through Lanes	Posted Speed
Crashes	1	0.82	0.28	-0.15	-0.16	-0.17	0.08	-0.01	0.19	0.01	-0.01	0.15	-0.13
Injuries	0.82	1	0.15	-0.14	-0.16	-0.16	0.07	-0.02	0.16	-0.003	0.0001	0.12	-0.13
Fatalities	0.28	0.15	1	-0.01	-0.02	-0.02	0.003	-0.01	0.05	0.04	-0.02	0.03	-0.01
15th Speed	-0.15	-0.14	-0.01	1	0.97	0.93	-0.77	-0.49	-0.47	0.55	-0.004	-0.18	0.76
Median Speed	-0.16	-0.16	-0.02	0.97	1	0.98	-0.7	-0.27	-0.46	0.54	-0.02	-0.17	0.81
85th Speed	-0.17	-0.16	-0.02	0.93	0.98	1	-0.54	-0.16	-0.41	0.51	-0.04	-0.12	0.85
85th - 50th	0.08	0.07	0.003	-0.77	-0.7	-0.54	1	0.57	0.46	-0.45	-0.05	0.29	-0.39
50th - 15th	-0.01	-0.02	-0.01	-0.49	-0.27	-0.16	0.57	1	0.22	-0.23	-0.05	0.1	-0.09
AADT	0.19	0.16	0.05	-0.47	-0.46	-0.41	0.46	0.22	1	-0.33	-0.17	0.55	-0.25
Length (mi)	0.01	-0.003	0.04	0.55	0.54	0.51	-0.45	-0.05	-0.33	1	0.08	-0.24	0.47
Func Class	-0.01	0.0001	-0.02	-0.004	-0.02	-0.04	-0.05	-0.05	-0.33	0.08	1	-0.17	-0.07
Through Lanes	0.15	0.12	0.03	-0.18	-0.17	-0.12	0.29	0.1	0.55	-0.24	-0.17	1	-0.04
Posted Speed	-0.13	-0.13	-0.01	0.76	0.81	0.85	-0.39	-0.09	-0.25	0.47	-0.07	-0.04	1

Chapter 5

Results

The results of two sets of models are presented in this section. First, models including absolute speed percentiles (15th, 50th, and 85th) are shown. Second, models including high and/or low speed differences (85th - 50th and 50th - 15th) are shown. Discussion of results can be found in the next section.

5.1. Speed percentiles and crash frequency

Table 6 shows the results of models relating speed percentiles to crashes on Georgia TMCs. The models shown in Table 6 chiefly differ in which percentile speeds are included. When all speeds are included (models 1 and 2), each percentile speed coefficient is negative, not significant and close to zero. However, when models do not include all three percentile speeds, sign and significance of each of the speed coefficients changes. When two are included (models 3-5), one or both speed coefficients become statistically significant. When only one percentile speed is included (models 6-8), its coefficient is negative and significant, with roughly the same magnitude, regardless of which speed measure is included.

5.2. Speed differences and crash frequency

Table 7 shows the model results of the speed differences model. When high speed difference is present, it is positive and significant in all model specifications. This result is robust to changes in the model, such as the inclusion of the low speed difference or controlling variables (models 1-3).

Low speed difference has a negative relationship with crashes that is much smaller in magnitude than high speed difference. In addition, when high speed difference is removed from the model, low speed difference is close to zero and not significant (model 4).

5.3. Other covariates

For all models of speed percentile and speed difference, the natural log of AADT, segment length, and the short TMC flag are positive and significant. The magnitude of AADT decreases substantially, however, when lane number, urban area, and functional class covariates are included. Posted speed limit is negative and significant in all cases. The coefficients for posted speed, segment length, and short TMCs do not change substantially based on which variables are included.

Urban area covariates are relative to the first urban area in the alphabet, Albany, GA. Rural areas, however, have a coefficient that is consistently negative and significant. Valdosta, GA is consistently significant and negative, with a larger magnitude of effect than other variables.

Since the vast majority of included TMCs are major arterials (Table 1), it is not surprising that the coefficients for minor arterials and collectors is not significant. However, local streets have positive and significant coefficients in most models (Table 6 and 7).

Table 6: Statewide Bike/Ped Crashes and Vehicle Percentile Speeds

	Crashes							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Constant	-6.865***	-2.085	-2.12	-1.999	-2.137	-2.429	-2.127	-1.95
15th Percentile Speed	0.001	-0.01		-0.024**	-0.007	-0.043***		
50th Percentile Speed	-0.069*	-0.034	-0.050**		-0.045**		-0.051***	
85th Percentile Speed	0.015	-0.009	-0.001	-0.033**				-0.056***
ln(AADT)	0.839***	0.430***	0.432***	0.434***	0.429***	0.442***	0.432***	0.464***
Miles	0.303***	0.320***	0.319***	0.318***	0.320***	0.318***	0.319***	0.309***
TMC < 0.5 Miles	-1.340***	-1.339***	-1.338***	-1.323***	-1.346***	-1.360***	-1.339***	-1.255***
Posted Speed (mph)	-0.033***	-0.032***	-0.032***	-0.030***	-0.033***	-0.044***	-0.033***	-0.029***
2 Lane Road		-0.969	-0.96	-0.992	-0.96	-0.974	-0.959	-1.01
3 Lane Road		-0.942	-0.942	-0.95	-0.936	-0.92	-0.941	-0.992
4 Lane Road		-0.344	-0.346	-0.346	-0.343	-0.354	-0.345	-0.373
5 Lane Road		-0.087	-0.099	-0.077	-0.085	-0.007	-0.098	-0.145
6 Lane Road		-0.035	-0.043	-0.024	-0.037	-0.011	-0.042	-0.054
7 Lane Road		0.078	0.057	0.109	0.072	0.164	0.057	0.05
8 Lane Road		-0.146	-0.161	-0.116	-0.153	-0.089	-0.161	-0.131
Athens		-0.376	-0.379	-0.362	-0.381	-0.372	-0.379	-0.337
Atlanta		-0.086	-0.085	-0.082	-0.087	-0.072	-0.085	-0.056
Augusta		-0.539	-0.535	-0.546	-0.535	-0.511	-0.535	-0.544
Brunswick		-0.768	-0.764	-0.775	-0.765	-0.767	-0.764	-0.773
Cartersville		-0.674	-0.683	-0.658	-0.679	-0.673	-0.683	-0.67
Chattanooga		0.5	0.496	0.514	0.496	0.532	0.496	0.536
Columbus		-0.588	-0.593	-0.578	-0.591	-0.569	-0.593	-0.577
Dalton		-0.607	-0.603	-0.609	-0.606	-0.589	-0.603	-0.585
Gainesville		-0.392	-0.394	-0.389	-0.391	-0.368	-0.394	-0.402
Hinesville		-0.502	-0.505	-0.517	-0.496	-0.488	-0.504	-0.586
Macon		-0.294	-0.303	-0.273	-0.301	-0.274	-0.304	-0.269
Rome		0.071	0.07	0.081	0.069	0.085	0.07	0.106
Rural Area		-0.788**	-0.801**	-0.776**	-0.790**	-0.756**	-0.800**	-0.829**
Savannah		0.157	0.155	0.157	0.16	0.216	0.156	0.145
Small Town		-0.491	-0.489	-0.489	-0.491	-0.495	-0.489	-0.475
Valdosta		-2.296**	-2.298**	-2.297**	-2.294**	-2.259**	-2.298**	-2.314**
Warner-Robins		-0.018	-0.015	-0.022	-0.016	-0.007	-0.015	-0.016
Minor Arterial		0.051	0.049	0.053	0.051	0.072	0.049	0.047
Collector		0.112	0.111	0.108	0.115	0.157	0.112	0.089
Local		1.149*	1.154*	1.141*	1.154*	1.171*	1.154*	1.147*
Observations	7,050	7,050	7,050	7,050	7,050	7,050	7,050	7,050
Log Likelihood	-2,599.14	-2,550.95	-2,551.16	-2,551.58	-2,551.02	-2,556.27	-2,551.16	-2,555.34
McFadden's R2	0.1392	0.1552	0.1551	0.155	0.1552	0.1534	0.1551	0.1537

Note:

*p<0.05; **p<0.01; ***p<0.001

Table 7: Statewide Bike/Ped Crashes and Vehicle Speed Distributions

	Crashes			
	(1)	(2)	(3)	(4)
Constant	-5.698***	-3.722**	-3.740**	-3.609**
High Speed Difference	0.089***	0.078***	0.066***	
Low Speed Difference	-0.021	-0.025		0.003
ln(AADT)	0.780***	0.540***	0.540***	0.588***
Miles	0.278***	0.297***	0.297***	0.282***
TMC < 0.5 Miles	-1.325***	-1.291***	-1.298***	-1.181***
Posted Speed (mph)	-0.081***	-0.075***	-0.077***	-0.077***
2 Lane Road	-1.145	-0.732	-0.769	-0.814
3 Lane Road	-1.306	-0.766	-0.785	-0.833
4 Lane Road	-0.666	-0.256	-0.279	-0.303
5 Lane Road	-0.295	0.111	0.13	0.105
6 Lane Road	-0.374	0.085	0.083	0.088
7 Lane Road	0.129	0.189	0.23	0.266
8 Lane Road	-0.413	0.02	0.039	0.112
Athens		-0.414	-0.394	-0.297
Atlanta		0.013	0.016	0.067
Augusta		-0.414	-0.412	-0.407
Brunswick		-0.719	-0.723	-0.739
Cartersville		-0.802	-0.789	-0.774
Chattanooga		0.592	0.609	0.68
Columbus		-0.574	-0.562	-0.548
Dalton		-0.469	-0.471	-0.428
Gainesville		-0.359	-0.346	-0.352
Hinesville		-0.583	-0.578	-0.69
Macon		-0.362	-0.333	-0.273
Rome		0.137	0.143	0.197
Rural Area		-0.901**	-0.869**	-0.911**
Savannah		0.314	0.333	0.346
Small Town		-0.466	-0.469	-0.446
Valdosta		-2.227**	-2.215**	-2.233**
Warner-Robins		0.056	0.058	0.064
Minor Arterial		0.098	0.107	0.112
Collector		0.226	0.236	0.223
Local		1.253*	1.257*	1.260*
Observations	7,050	7,050	7,050	7,050
Log Likelihood	-2,611.49	-2,574.85	-2,576.32	-2,584.90
McFadden's R2	0.1351	0.1473	0.1468	0.1439
Note:	*p<0.05; **p<0.01; ***p<0.001			

Chapter 6

Discussion

The results of this study demonstrated some important real-world implications for how speed is measured in reference to safety of bicyclists and pedestrians. First and foremost, differences in speed may be more effective in assessing safety than more commonly-used measures, like the 85th percentile speed. In what follows, some conclusions are drawn regarding the real-world impact of the results. Then, a few limitations of the study are mentioned. Finally, the implications for transportation planning practices are explored.

6.1. Speed differences

Table 7 shows that the high speed difference on the roadway is significantly related to bicycle and pedestrian crash frequency, regardless of the model specification. This finding is consistent with Ederer et al. (2020) and provides evidence that the spread of the high end of the vehicle speed distribution has an effect on crash likelihood with bicycles and pedestrians. The more vehicles exceed the prevailing speed of traffic, the more they induce conflict with other vehicles and active road users, with less and less time to react and avoid conflicts as they develop.

Furthermore, the coefficient for low speed difference has low significance and magnitude in each of the models in which it is included. This finding confirms that changes in vehicle speeds on the high end of the speed distribution have a greater effect on the safety of bicyclists and pedestrians than changes in lower speeds. So, measures of spread which weigh both ends of

the speed distribution equally, like standard deviation, are not appropriate for predicting crashes involving bicyclists and pedestrians.

Over the course of one year, there may be a number of reasons why a TMC has a large spread of vehicle speeds at the high end of the distribution. Arterial roadway segments with high free flow speeds but consistent congestion patterns might have a greater high speed difference. Alternatively, roadways with high intersection density or many driveways providing access to busy commercial space might be characterized by a high volume of merging vehicles which enter the roadway at significantly reduced speeds. In each of these cases, greater differences in speed (especially faster vehicles at the high end) create greater chances for conflict with more serious implications in the event of a crash.

Each of these possible causes highlights a scenario with high complexity and high mental workload for all road users as they react and adjust to the roadway environment. This translates to less attention paid to pedestrians and bicyclists. Furthermore, in moments of conflict, drivers reacting defensively might by reflex avoid conflicting vehicles first, which pose the greatest danger to the drivers themselves. Bicyclists and pedestrians may go unnoticed or simply be noticed later. While there are many different possible scenarios, one thing is certain--if the fastest vehicles travel even faster in these complex situations, there is less reaction time and less room for error. Braking distances must increase, small mistakes may compound, and chances of conflict may also increase. Finally, in the event of a crash, higher speeds increase the risk of severe or fatal injury.

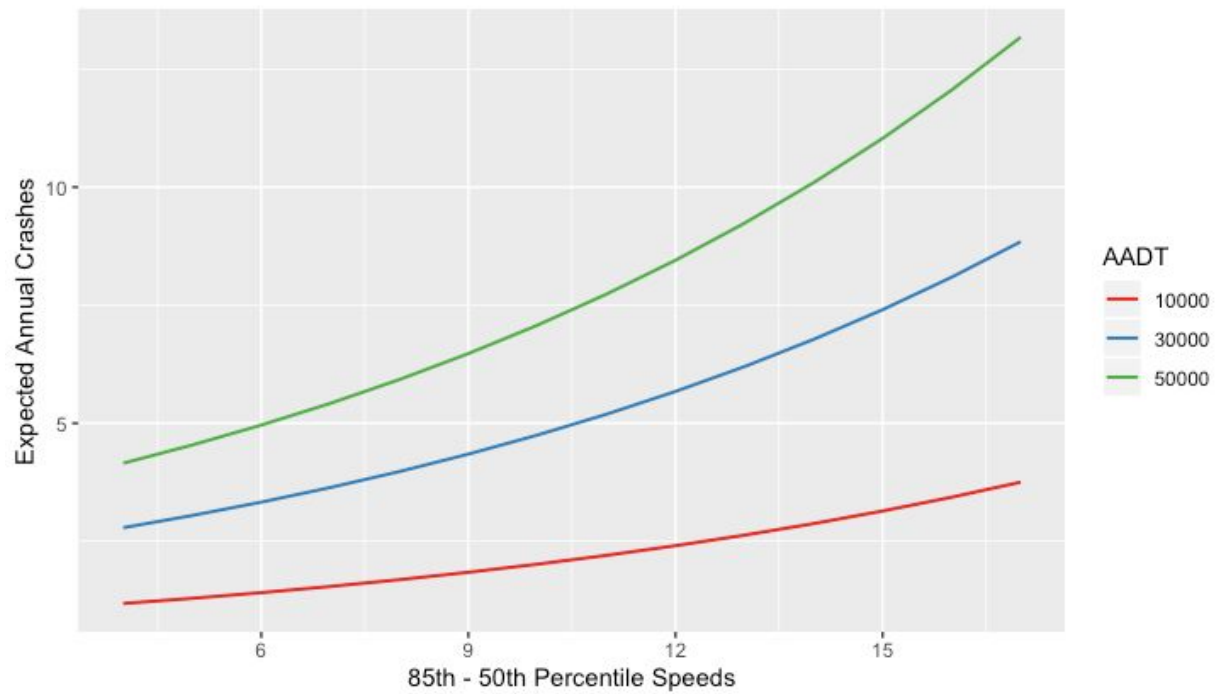


Figure 15: Estimated marginal effects of 85th - 50th percentile speeds on crashes per 10 miles

Since Negative Binomial coefficients are difficult to interpret on their own, examining their marginal effects is helpful for understanding their importance. Figure 15 shows the modeled marginal effect of changes in high speed difference on 10 miles of Georgia roadway. The effect increases as high speed difference increases, and it varies in magnitude by AADT. For the highest speed difference, highest volume roads, a 5 mph reduction in 85th - median speed can result in as much as a reduction of 5 annual crashes per 10 miles. Since the number of crashes included in this study is relatively small, the confidence bands for these marginal effects curves are quite large and overlapping. With a limited number of crashes, we cannot confirm that these curves are precise as shown. A broader bicycle and pedestrian study should be completed to confirm these results.

6.2. Speed percentiles

As outlined in the Results section, the coefficients for speed 15th, 50th and 85th percentiles change in sign, magnitude and significance when different combinations of the three are included (Table 6). As mentioned before and shown in Table 5, these speed percentiles are each highly correlated with the other. This suggests a multicollinearity concern when more than one speed percentile is included. This result has statistical and intuitive implications for the speed percentile model. Highly collinear covariates significantly affect the estimation of variance in the coefficient. This can be seen in models 1 and 2 of Table 6 above, where the percentile coefficient estimates are generally not statistically significant. It is possible that multicollinearity has inflated the variance of the coefficient estimates. Using more than one speed percentile to describe the speed distribution in these models yields inconclusive results.

Regardless of these results, speed percentiles may not be able to provide an intuitive measure of speed distribution. The coefficients in the models in Table 6 are estimates of the change in crashes due to a one unit change in the explanatory variable, holding other variables constant. A unit increase in the 50th percentile speed, for example, would result in a unit decrease of the high speed difference. Considering this fact and the high correlation among speed percentiles (Table 5), it may not make practical sense to consider adjusting any single percentile speed without the other speeds in the distribution changing as well. While speed percentiles together may describe the distribution, their presence together in a statistical model defies simple interpretation.

However, when any one percentile speed is included, its coefficient is negative and significant (Table 6. Models 6-8). These model results show how using single speed percentiles

can yield counterintuitive or even confusing results. As noted by Baryua (1998), the negative relationship between speed and crashes could be confounded by conditions of congestion. Over the course of a year, greater periodic congestion may reduce the 85th percentile speed by adding very low speeds to the distribution. As seen in the subsection above, congestion also leads to environments of greater complexity and potential for conflict. Without information on congestion, one cannot draw a clear conclusion regarding the relationship between any one speed percentile and bicyclist and pedestrian crash frequency.

6.3. Posted speed

As mentioned in the results section, posted speed has a consistently negative and significant relationship with crash frequency in both the speed difference and speed percentile models. It is unlikely that higher speed limits lead to safer conditions for bicyclists and pedestrians; rather, the speed limit may be capturing some of the effect of roadway context in these models. Speed limits are likely higher on rural arterials than on urban arterials, where more bicyclists and pedestrians are present. So, crashes on higher speed limit arterials are less common because bicyclist and pedestrian exposure is lower on those roadways. Bicyclist and pedestrian count data are necessary to fully understand the role of speed limit, and vehicle speeds, in studies like these.

In the speed difference models (Table 7), the coefficient for posted speed has a much greater magnitude compared to the speed percentile models (Table 6). This finding demonstrates that posted speed likely captures the effect of location of the speed distribution. In the speed percentile models, this effect may be shared between a percentile operating speed and the posted

speed. This finding is consistent with literature that shows a close relationship between operating and posted speed (Fitzpatrick et al. 2001). Posted speed and speed difference should be included together in speed-safety models as a way to represent the absolute speed and the speed distribution, respectively. These two metrics provide a relatively simple, intuitive and complete description of the speed profile of a roadway.

Posted speed is included here as a numeric variable with a linear relationship to crash frequency. While the effect of posted speed is clear in the models, it is not clear that this representation of posted speed is the most appropriate. Further study is needed to apply posted speed in its most sensible form.

6.4 Other covariates

For the most part, coefficients for urban area dummy variables were not statistically significant. One important exception is the coefficient for rural areas relative to Albany, GA, which was consistently negative in all models. This result makes intuitive sense, given that bicycle and pedestrian activity is lower in less dense areas, and exposure to a crash is thus reduced. Further study should be conducted to understand whether the consistently large negative coefficient in Valdosta may indicate some other unobserved characteristic. The coefficients for urban areas describe underlying heterogeneity within each urban area. The coefficient for Atlanta, for example, is not significant and close to zero, which indicates that the effects unique to Albany, GA (the base case) are not different from those in Atlanta. Atlanta has a large variety of community contexts and roadway environments, as does Albany and each other urban area. Their population density also ranges significantly depending on the particular locale.

All of these competing effects seem to drop out in this analysis. Further studies should include a more granular breakdown of urban versus rural conditions.

TMC length is positive and significant in all models. This result makes intuitive sense, given that longer TMCs have more area for a crash to occur. Exposure increases with TMC length, so crash frequency might follow as well. However, the flag for short TMCs consistently has a negative and significant coefficient. As mentioned in the Data and Methodology sections, short TMCs are frequently internal TMCs, or short links between access points from intersecting or flyover roadways with ramps or intersections with channelized right turns. In these cases, conflict from merging vehicles may occur more often just downstream from the internal TMCs, since the merge point marks the end of the internal TMC.

6.5. Limitations

Data restrictions and the network scope of the analysis prevent the inclusion of important factors like variations in weather and lighting at the time of the crash, sociodemographic characteristics and detailed roadway characteristics, which have been shown to play a factor in crashes when studied at the corridor or other micro level. These factors, along with granular population and demographic data, require an extremely data-intensive approach to implement at this scale, and attributing these factors would be a significant challenge for a study area comparable to the Georgia arterial network. Nevertheless, future work would improve on this study by incorporating these more granular factors.

The most important information missing from this study is bicyclist and pedestrian count data. Active transportation activity is likely confounding the effect of posted speed and percentile

speed, since places with higher speeds are places where people are less likely to walk and bike. It is also possible that the effect of high speed difference would change given this information. Including count information would clarify the interpretation of the model overall.

6.6. Policy implications

This study reveals that transportation planning professionals must expand its data collection efforts and reprioritize roadway space for the sake of the safety of active transportation users. The recommendations below amount to a long-overdue sea change in the priorities of transportation planners, from minimizing vehicle travel times to ensuring road user safety.

First, a comprehensive and coordinated effort to collect bicyclist and pedestrian counts is critical to understanding bicyclist and pedestrian safety needs. This study has shown that measuring crash risk for these road users is much more effective with count data. Currently, decisions to install bicycle and pedestrian infrastructure are not consistently based on the true levels of risk and exposure on roadways, because that information is not consistently available. Georgia DOT and many other state agencies do not have robust data collection programs, and county and municipal governments cannot reasonably build a cohesive database without state coordination and funding. Furthermore, without state-level information, state DOTs are unlikely to enact policies at the state level. Just as U.S. states collect vehicle count data, bicyclist and pedestrian counts should be prioritized to effectively measure crash risk and improve further studies of vehicle speed and active transportation safety.

In addition to collecting count data, state agencies must keep a comprehensive record of sidewalks, crossings, and other infrastructure for active transportation, like bike lanes, cycle

tracks, and greenways. These infrastructure improvements have proven to reduce crash risk (Teschke et al. 2012), and their records are generally kept by local governments, if at all. States ought to have a complete record for several reasons. First, this information is necessary across municipalities to create a plan for a cohesive network. Second, if this database existed, states could include information on infrastructure in broad studies and gain greater understanding as to the impact of these kinds of infrastructure on bicyclist and pedestrian safety.

With count and infrastructure data available, state DOTs and local governments can effectively identify road segments with high crash risk to pedestrians and bicyclists, and implement design changes to improve their safety. Planners should search for areas that not only have a crash history, but have a risk of future crashes--places with high speed limits, many active transportation users, and large speed differences. These areas have high exposure for bicyclists and pedestrians in combination with a risky vehicle speed profile.

Once identified, there are many possible solutions for improving the safety of bicyclists and pedestrians, and each set of solutions should be based on the particular context of the site. This research suggests that, in all cases where the vehicle speed profile is a concern, planners and engineers should work to reduce the speed of the fastest vehicles on the roadway.

Principally, planners should recognize that many arterial roadways simultaneously attempt to serve two competing goals-- access and mobility. Mobility is currently prioritized through high speed limits, large rights-of-way with many lanes, and long light cycles. Vehicle access is currently prioritized with frequent driveways, often directly leading to each parcel of land. These aspects of the roadway function at odds in many ways, and they result in a high-speed environment with frequent points of conflict. Frequent driveways are also a danger

for pedestrians; they break up the sidewalk with many potential points of conflict where drivers are rightfully concerned about entering or exiting a high speed traffic flow. These complex environments, described above, are characterized by high speed differences. They are a safety concern for all road users.

So, planners must prioritize either access or mobility on arterial roadways. These decisions should be driven by a combination of considerations, especially current usage by all road users, surrounding land uses and population density, and creating an effective and connected network for vehicles, bicycles and pedestrians.

For mobility-oriented roads, a priority should be consolidating and signaling vehicle access points. This treatment reduces the number of potential conflicts among vehicles as they enter the roadway, and keeps more of the sidewalk safe and continuous for pedestrians. In addition, transportation professionals should separate bicyclists and pedestrians with dedicated infrastructure (through grade, physical barriers, and/or significant space). All of these treatments can reduce conflicts among vehicles, tighten the distribution of vehicle speeds, and reduce conflicts for bicyclists and pedestrians.

To prioritize access on arterial roadways, transportation professionals should focus on reduced speed limits, reduced vehicle right-of-way, short light cycles, and sufficient space for pedestrians, bicyclists and transit. Narrower lanes and treatments like street trees, which narrow the driver's field of vision, will naturally enforce reasonable vehicle speeds while shielding pedestrians from the vehicle travelway. These treatments should offer sufficient reaction time for all road users to avoid a crash, and reduce severity in the event of a crash. Importantly, these treatments also reduce the fastest speeds, potentially reducing the high end of the vehicle speed

distribution. Planners, engineers and policymakers should be sure to consider context when planning these roadways.

As development continues to occur on arterial roadways, planners and developers should keep in mind the “3 D’s”: density, design and diversity (Cervero & Kockelman, 1997). In urban areas, mixed land uses, pedestrian scaled storefronts, and spaces oriented towards the street, not just the parking lot, should be prioritized (Cervero & Kockelman 1997, Dumbaugh & Rae, 2009). In all cases, shared parking in the rear of the lot, with few access points, minimizes conflict between vehicles and pedestrians, reduces necessary space for cars, and engages the space for all road users. These changes should be incorporated into both zoning ordinances and development practices.

On any arterial roadway, one important choice is setting the posted speed limit. First, the common practice of setting the posted speed at the 85th percentile speed on arterials (NTSB 2017) should end. This practice normalizes the higher end of the speed distribution and is likely to encourage greater differences between the median speed and the fastest vehicles. Second, speed limits should be set based on the full context of the roadway, and the prevailing speed of vehicles is only one small part of that context, especially in areas with significant bicycle and pedestrian activity. Research has shown that in crashes involving a pedestrian with impact speeds of 40 mph or greater, the pedestrian is likely to sustain a severe injury or fatality (Tefft 2013). Crash risk decreases significantly as speed decreases. In these contexts, speed limits greater than 35 mph amount to an acceptance of injuries and deaths for unprotected users. While it is beyond the scope of this study to determine specific speed limits for particular contexts, for access-oriented arterials, speed limits of 40 and 45 mph are too high.

One frequent challenge in speed limit reductions is ensuring enforcement. Despite its controversial history in the United States, automated speed enforcement (ASE) has proven to be effective in improving speed adherence, especially among the fastest speeders (NTSB 2017). ASE is an important tool in reducing the fastest speeds in the speed distribution. ASE should be implemented on roadways with speeding concerns and those with significant bicycle and pedestrian activity as a way to protect these road users from a potentially severe crash.

Significant changes in the approach to planning and design for arterial roadways are needed to prioritize the safety of drivers, bicyclists and pedestrians. These changes must prioritize the goals of reductions in vehicle speeds overall, tightening the high end of vehicle speed distributions, and reducing exposure for bicyclists and pedestrians. These goals can only be effectively achieved if states embark on comprehensive bicycle and pedestrian count data collection efforts.

6.7. Summary

The results of this study show that the difference in 85th and 50th percentile vehicle speeds is a robust indicator of crash risk for bicyclists and pedestrians on the Georgia arterial network. Furthermore, the high speed difference has a more significant relationship with crash risk than does the low speed difference. So, measures of the whole distribution like standard deviation are not as appropriate as metrics emphasizing the high end of the distribution. While some of the covariates included could be improved with greater granularity, their sign and significance are consistent with previous literature. TMC length, AADT, and posted speed are the critical roadway characteristics to include at this large scale of analysis.

Transportation professionals must focus on reducing the high end of speed distributions, reducing speeds overall, and protecting bicyclists and pedestrians from conflict with vehicles. An important step is collecting count data for these road users. Additionally, planners should define arterials as access- or mobility-oriented, and update roadway environments accordingly.

Chapter 7

Conclusion

This study evaluated the relationship between vehicle speeds and frequency of crashes involving bicyclists or pedestrians on a large portion of the network of Georgia arterial roadways. It is the first network-level active transportation safety study of its kind to use probe vehicle speed data on a large scale. Many area crash studies imply that vehicle speeds are an important factor in crashes, and this study provides further supporting evidence. Vehicle speed distributions, especially their high end, are a contributing factor for the crash risk of bicyclists and pedestrians on arterial roadways.

Negative Binomial models were used to evaluate the impact of speed differences and percentile speeds on bicycle and pedestrian safety. As hypothesized, the difference between the 85th and 50th percentile speeds is a robust indicator of bicycle and pedestrian crashes. These results indicate that complex arterial environments with intermittent congestion or high volumes of vehicle access are more and more dangerous as the fastest drivers increase their speeds. Bicyclists and pedestrians are both the last to be noticed in moments of conflict, and they are most likely to be severely or fatally injured.

Speed percentile models, when included individually, were negatively related to crash frequency. However, when included together, the sign, significance and magnitude of their coefficients changed depending on the model specification. In contrast, speed differences were robust to changes in models. Speed percentiles are more difficult to interpret and at least partially describe the level of operating speed. Posted speed effectively describes the level of operating

speed on its own. Speed differences have proven to more effectively and intuitively describe the distribution of speeds than speed percentiles.

Additionally, the low speed difference (50th - 15th percentile speeds) was shown to have a weaker relationship with crash frequency than the high speed difference. This result reinforces the notion that all parts of the speed distribution are not equally significant in predicting crashes with bicyclists and pedestrians. Measures of speed spread that consider the whole distribution, like standard deviation, are not appropriate in their application to bicycle and pedestrian safety. More metrics that heavily weigh the high end of the speed distribution should be tested.

While this study has shown clear and intuitive results, it could be developed further for a better understanding of the speed and safety relationship. Due to their scope and scale, network level studies such as these require careful choice of methods and included covariates.

First and foremost, more granular measures of population, demographics, and active transportation activity should be included. Controlling for bicycle and pedestrian activity--which is a key measure of exposure to crashes for vulnerable road users--ensures that other variables like posted speed and functional class are not expressing some of the effect of transportation activity levels. Demographic effects, especially race and income, should be considered as well.

However, fundamental limitations exist with TAZs and census tracts in reference to activity on arterial roadways. Roadways with high functional class are commonly used as boundaries for these territories, especially major arterials, so crash hotspots could be arbitrarily split between several territories (Xie et al 2017). Furthermore, spatial association of territories to arterial road segments on their boundaries would be highly dependent on arbitrary quirks of the spatial layout of the two data sets (such as coordinate reference systems and truncation of lat/lon

values). This challenge associating demographic information to arterials proves to be an ongoing limitation of spatial crash analysis at the macroscopic level.

Another challenge to incorporating new information is the lack of available bicycle and pedestrian count data on a broad scale. Network level crash studies will not be able to reliably predict crash risk for bicyclists and pedestrians without these data inputs.

Developing an understanding of congestion is also critical to isolating the effect of speed variation to crashes with bicyclists and pedestrians. Currently, the high speed difference is capturing some of the effects of congestion; as congestion accumulates and dissipates, those differences in speed create conflict and complexity that likely results in bicycle and pedestrian crashes. Speed percentiles capture congestion as well, but mask its effect among the majority of roadways which rarely experience congestion. Considering congestion and speed differential separately may give more insight as to how arterial roadways can be designed for tighter speed distributions and greater safety.

Finally, accounting for spatial random effects through Bayesian or other spatial methods would improve model fit and account for unobserved effects in the models.

The implications of this study are significant to the practices of transportation professionals. Planners, engineers and policymakers have an important responsibility to act to prevent injuries and deaths among active transportation users. First and foremost, more data is needed on pedestrian and bicycle counts statewide to make informed decisions. Data collection efforts at the state level are critical. Then, arterials should prioritize vehicle access or mobility, not both, to create safe and connected networks for all road users. Design of roadways and surrounding land uses should reflect these priorities. Finally, significant changes in the practices

of setting speed limits are needed. Speed limits should be set based on context and should prioritize crash severity reduction in contexts with many unprotected users. The widespread practice of setting vehicle speeds based on the 85th percentile speed should end. Finally, states and municipalities should transparently embrace automated speed enforcement. Each of these recommendations focuses on reducing exposure for bicyclists and pedestrians, reducing the overall level of operating speed, and tightening the distribution of speeds at the high end. These three goals ought to be foremost in transportation planning for pedestrian and bicyclist safety.

As the nation continues to urbanize, safety of bicyclists and pedestrians is growing more and more central to the responsibilities of transportation professionals across the United States. Intuitively, we know that speed kills, and these deaths are both preventable and costly to society. It is time to make changes to the way we consider vehicle speeds in our operational decisions. Doing so will yield dividends in terms of economy, public health, and social justice.

References

- Aarts, L. & van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention*, 38, 215-224. <https://doi.org/10.1016/j.aap.2005.07.004>
- AASHTO (2010). *The Highway Safety Manual*, American Association of State Highway Transportation Professionals, Washington, D.C., <http://www.highwaysafetymanual.org>
- Aguero-Valverde, J., & Jovanis, P. P. (2008). Analysis of Road Crash Frequency with Spatial Models. *Transportation Research Record*, 2061(1), 55–63. <https://doi.org/10.3141/2061-07>
- Baruya, A. (1998). Speed-Accident Relationships on European Roads. *Road Safety in Europe: 9th International Conference*. Bergisch Gladbach, Germany. <http://virtual.vtt.fi/virtual/proj6/master/pre11.pdf>
- Cervero, R. & Kockelman, K. (1997). Travel Demand and the 3Ds: Density, Diversity and Design. *Transportation Research Part D: Transport and Environment*, 2 (3). [https://doi.org/10.1016/S1361-9209\(97\)00009-6](https://doi.org/10.1016/S1361-9209(97)00009-6)
- Chen, P. (2015). Built environment factors in explaining the automobile-involved bicycle crash frequencies: A spatial statistic approach. *Safety Science*, 79, 336 - 343. <https://doi.org/10.1016/j.ssci.2015.06.016>
- Chen, P. & Zhou, J. (2016). Effects of the built environment on automobile-involved pedestrian crash frequency and risk. *Journal of Transport & Health*, 3(4). <https://doi.org/10.1016/j.jth.2016.06.008>
- Hu, W. & Cicchino, J. (2018). An examination of the increases in pedestrian motor-vehicle crash fatalities during 2009–2016. *Journal of Safety Research*, 67, 37-44. <https://doi.org/10.1016/j.jsr.2018.09.009>
- Dumbaugh, E. & Rae, R. (2009). Safe Urban Form: Revisiting the Relationship Between Community Design and Traffic Safety. *Journal of the American Planning Association*, 75 – 3. <https://doi.org/10.1080/01944360902950349>
- Ederer, D., Rodgers, M., Hunter, M., & Watkins, K. (2020). Probe-speed based safety performance metrics in Georgia: A case study. Presented at the 99th Annual Meeting of the Transportation Research Board.

- Edquist, K., Rudin-Brown, C. & Lenne, M. (2012). The effects of on-street parking and road environment visual complexity on travel speed and reaction time. *Accident Analysis & Prevention*, 45, 759-765. <https://doi.org/10.1016/j.aap.2011.10.001>
- Elvik, R., Christensen, P., & Amundsen, A.H. (2004). Speed and road accidents: an evaluation of the power model. Institute for Transport Economics, 740. <https://www.toi.no/getfile.php?mmfileid=1007>
- Elvik, R., Vadeby, A., Hels, T., & van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accident Analysis & Prevention*, 123, 114-122. <https://doi.org/10.1016/j.aap.2018.11.014>
- Fitzpatrick, K., Carlson, P., Brewer, M., & Wooldridge, M. (2001). Design factors that affect driver speed on suburban streets. *Transportation Research Record*, 1751(1), 18-25. <https://doi.org/10.3141/1751-03>
- Geyer, J., Raford, N., Pham, T., & Ragland, D. R. (2006). Safety in Numbers: Data from Oakland, California. *Transportation Research Record*, 1982, 150–154. <https://doi.org/10.1177/0361198106198200119>
- Hauer, E., Harwood, D., Council, F. & Griffith, M. (2002). Estimating Safety by the Empirical Bayes Method. *Transportation Research Record*, 1784, 126-131. <https://doi.org/10.3141/1784-16>
- Insurance Institute for Highway Safety. (2019). Fatality Facts 2018: Bicyclists. <https://www.iihs.org/topics/fatality-statistics/detail/bicyclists>
- Insurance Institute for Highway Safety. (2019). Fatality Facts 2018: Yearly Snapshot. <https://www.iihs.org/topics/fatality-statistics/detail/yearly-snapshot>
- Kaplan, S. & Prato, C. (2015). A Spatial Analysis of Land Use and Network Effects on Frequency and Severity of Cyclist–Motorist Crashes in the Copenhagen Region. *Traffic Injury Prevention*, 16(7), 724-731. <https://doi.org/10.1080/15389588.2014.1003818>
- Kroyer, H., Jonsson, T. & Varhelyi, A. (2014). Relative fatality risk curve to describe the effect of change in the impact speed on fatality risk of pedestrians struck by a motor vehicle. *Accident Analysis & Prevention*, 62, 143-152. <https://doi.org/10.1016/j.aap.2013.09.007>

- Lord, D., Washington, P., & Ivan, J. (2005). Poisson, Poisson-gamma and Zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention*, 37, 35-46. <https://doi.org/10.1016/j.aap.2004.02.004>
- Ma, M., Yan, X., Abdel-Aty, M., Huang, H. & Wang, X. (2010). Safety Analysis of Urban Arterials Under Mixed-Traffic Patterns in Beijing. *Transportation Research Record*, 2193, 105-115. <http://dx.doi.org/10.3141/2193-13>
- National Highway Transportation Safety Administration. (2017). Traffic Safety Facts 2015: Pedestrians. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812375>
- National Highway Transportation Safety Administration. (2017). Traffic Safety Facts 2015: Bicyclists and Other Cyclists. <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812382>
- Noland, R. & Quddus, M. (2005). Congestion and safety: A spatial analysis of London. *Transportation Research Part A: Policy and Practice*, 39(7-9), 737-754. <https://doi.org/10.1016/j.tra.2005.02.022>
- Noland, R., Quddus, M., Ochieng, W. (2008). The Effect of the London Congestion Charge on Road Casualties: An Intervention Analysis. *Transportation*, 35, 73-91. <https://doi.org/10.1007/s11116-007-9133-9>
- National Performance Management Research Data Set Analytics. (N.D.). NPMRDS FAQs. <https://nprmds.ritis.org/analytics/help/#tmcs>
- Oikawa, S., Matsui, Y., Doi, T., Sakurai, T. (2016). Relation between vehicle travel velocity and pedestrian injury risk in different age groups for the design of a pedestrian detection system. *Safety Science*, 82, 361-367. <https://doi.org/10.1016/j.ssci.2015.10.003>
- Parker, M. R. (1997) Effects of Raising and Lowering Speed Limits on Selected Roadway Sections (FHWA-RD-92-084). Federal Highway Administration U.S. Department of Transportation. <https://www.fhwa.dot.gov/publications/research/safety/97084/97084.pdf>
- Pirdavani, A., Daniels, S., van Vlieden, K., Brijs, K. & Kochan, B. (2017). Socioeconomic and sociodemographic inequalities and their association with road traffic injuries. *Journal of Transport & Health*, 4, 152-161. <https://doi.org/10.1016/j.jth.2016.12.001>

- Quddus, M. (2008). Modelling area-wide count outcomes with spatial correlation and heterogeneity: An analysis of London crash data. *Accident Analysis & Prevention*, 40(14), 1486-1497. <https://doi.org/10.1016/j.aap.2008.03.009>
- Rosen, E. & Sander, U. (2009). Pedestrian Fatality Risk as a Function of Car Impact Speed. *Accident Analysis & Prevention*, 41(3), 536-542. <https://doi.org/10.1016/j.aap.2009.02.002>
- Saha, D., Alluri, P. & Gan, A. (2015). Prioritizing Highway Safety Manual's Crash Prediction Variables Using Boosted Regression Trees. *Accident Analysis & Prevention*, 79, 133-144. <https://doi.org/10.1016/j.aap.2015.03.011>
- Siddiqui, C., Abdel-Aty, M. & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, 45, 382-391. <https://doi.org/10.1016/j.aap.2011.08.003>
- Silvano, A., & Bang, K. (2016). Impact of Speed Limits and Road Characteristics on Free-Flow Speed in Urban Areas. *Journal of Transportation Engineering*, 142(2). [http://dx.doi.org/10.1061/\(ASCE\)TE.1943-5436.0000800](http://dx.doi.org/10.1061/(ASCE)TE.1943-5436.0000800)
- Song, T., So, J., Lee, J. & Williams, B. (2017). Exploring Vehicle–Pedestrian Crash Severity Factors on the Basis of In-Car Black Box Recording Data. *Transportation Research Record*, 2659, 148-154. <https://doi.org/10.3141/2659-16>
- Solomon, D. (1964). Crashes on main rural highways related to speed, driver and vehicle. Bureau of Public Roads. U.S. Department of Commerce. United States Government Printing Office, Washington, D.C.
- Strauss, J., Miranda-Moreno, L. & Morency, P. (2015). Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts. *Accident Analysis & Prevention*, 83, 132-142. <https://doi.org/10.1016/j.aap.2015.07.014>
- Sze, N. & Wong, S. (2007). Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accident Analysis & Prevention*, 39(6), 1267-1278. <https://doi.org/10.1016/j.aap.2007.03.017>
- Tarko, A & Azam, S. (2011) Pedestrian injury analysis with consideration of selectivity bias in linked police-hospital data. *Accident Analysis & Prevention*. 49(5), 1689-1695. <https://doi.org/10.1016/j.aap.2011.03.027>

- Tasic, I., Elvik, R. & Brewer, S. (2017). Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. *Accident Analysis & Prevention*, 109, 36-46. <https://doi.org/10.1016/j.aap.2017.07.029>
- Taylor, M.C., Lynam, D.A., & Baruya, A. (2000). The effects of drivers' speed on the frequency of road accidents (ISSN 0968-4107). Transport Research Laboratory Report No. 421. <https://trl.co.uk/reports/TRL421>
- Tefft, B. C. (2013). Impact speed and a pedestrian's risk of severe injury or death. *Accident Analysis & Prevention*, 50, 871-878. <https://doi.org/10.1016/j.aap.2012.07.022>
- Teschke, K. et al. (2012). Route Infrastructure and the Risk of Injuries to Bicyclists: A Case-Crossover Study *American Journal of Public Health*, 102, 2336-2343, <https://doi.org/10.2105/AJPH.2012.300762>
- U.S. Environmental Protection Agency. (2018). Ecoregions. <https://www.epa.gov/eco-research/ecoregions>
- Ukkusuri, S., et al. (2012). The role of built environment on pedestrian crash frequency. *Safety Science*, 50(4), 1141-1151. <https://doi.org/10.1016/j.ssci.2011.09.012>
- Vandenbulcke, G., Thomas, I., Panis, L. (2014). Predicting cycling accident risk in Brussels: A spatial case-control approach. *Accident Analysis & Prevention*, 62, 341-357. <https://doi.org/10.1016/j.aap.2013.07.001>
- Wang, Y. & Kockelman, K. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention*, 60, 71-84. <https://doi.org/10.1016/j.aap.2013.07.030>
- Wang, X., Zhou, Q., Quddus, M., Fan, T., & Fang, S. (2018). Speed, speed variation and crash relationships for urban arterials. *Accident Analysis and Prevention*, 113, 236-243. <https://doi.org/10.1016/j.aap.2018.01.032>
- Xie, K., Ozbay, K., Kurkcu, A. & Yang, H. (2017). Analysis of Traffic Crashes Involving Pedestrians Using Big Data: Investigation of Contributing Factors and Identification of Hotspots. *Risk Analysis*, 37(8), 1459-1476. <https://doi.org/10.1111/risa.12785>
- Xin, C., Guo, R., Wang, Z., Lu, Q. & Lin, P. (2017). The effects of neighborhood characteristics and the built environment on pedestrian injury severity: A random parameters generalized

ordered probability model with heterogeneity in means and variances. *Analytic Methods in Accident Research*, 16, 117-132. <https://doi.org/10.1016/j.amar.2017.10.001>

Zahabi S., Strauss, J., Manaugh, K., and Miranda-Moreno, L. (2011). Estimating Potential Effect of Speed Limits, Built Environment, and Other Factors on Severity of Pedestrian and Cyclist Injuries in Crashes. *Transportation Research Record*, 2247, 81-90.
<https://doi.org/10.3141/2247-10>